

AFIT/GEE/ENV/95D-19

AN ANALYSIS OF  
1994 AIR FORCE  
TOXIC RELEASE INVENTORY  
REPORTING

THESIS

Brian A. Pollock, Captain, USAF

AFIT/GEE/ENV/95D-19

19960205 048

Approved for public release; distribution unlimited

EXEMPT FROM AUTOMATIC DOWNGRADING AND DECLASSIFICATION

The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

AN ANALYSIS OF  
1994 AIR FORCE  
TOXIC RELEASE INVENTORY  
REPORTING

THESIS

Brian A. Pollock, B.S.I.E.  
Captain, USAF

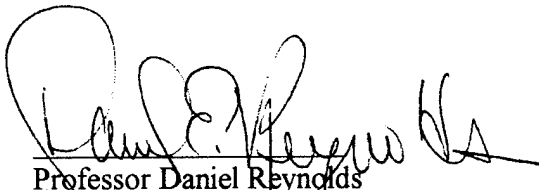
Presented to the Faculty of the School of Engineering

of the Air Force Institute of Technology

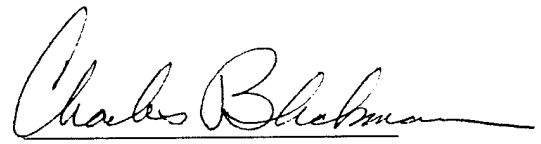
In Partial Fulfillment of the

Requirements for the Degree of

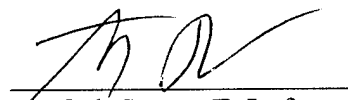
Master of Science in Engineering and Environmental Management



Professor Daniel Reynolds  
Reader



Dr. Charles A. Bleckmann  
Reader



Lt. Col. Steven T. Lofgren  
Thesis Advisor

### Acknowledgments

As with all things in my life, I must first give thanks to my Lord and Savior, Jesus Christ, without whom I would accomplish nothing of value. I also wish to acknowledge the members of my thesis committee. Lt Col Steve Lofgren saw me from the genesis of an idea through the difficulties associated with making my effort both academically and practically valuable. His genuine concern for my effort and the fact that he is a fellow Industrial Engineer helped make my research a pleasure. Dr. Charles Bleckmann's experience with the thesis process kept the rhythm of the thesis process at a steady beat. Professor Dan Reynolds' enthusiasm, statistical expertise and constant good humor kept overwhelming frustration at bay when I could have least afforded to succumb to it. A special word of appreciation goes to Capt. Dave Nelson who sponsored this research and provided a large chunk of the data. This work was due to his inspiration and I hope it contributes to the continuous process improvement of the Air Force's reporting of Toxic Release Inventory data. Last, but certainly not least, I thank my family: my wife, Joanie and sons Tyler and Zachary. Together, they are a constant reminder that while my research may push the frontier of knowledge back a little, my most important and lasting contributions to this world will be a result of my role as a husband and father.

Brian A. Pollock

## Table of Contents

	Page
Acknowledgments .....	ii
List of Figures .....	v
List of Tables .....	vi
Abstract .....	vii
I. Introduction .....	1
General Issue .....	2
Problem Statement .....	2
Research Objectives .....	4
Scope of Research .....	5
II. Literature Review .....	6
EPCRA (SARA Title III) .....	7
Executive Order 12856 .....	11
Air Force Environmental Guidance .....	11
Release Estimating Literature .....	12
EPA Guidance .....	12
Industry Approaches .....	12
DoD Guidance .....	13
Air Force Guidance .....	14
GAIA Reference Materials .....	18
III. Methodology .....	20
Research Objective 1 .....	20
Data Collection .....	20
Obtaining and Organizing Data .....	21
Statistical Analysis .....	21
Research Objective 2 .....	22
Data Collection .....	22
Obtaining and Organizing Data .....	26
Statistical Analysis .....	27
IV. Findings and Analysis .....	32
Results of Statistical Analysis .....	32
Research Objective 1 .....	32
Research Objective 2 .....	40

V. Conclusions and Recommendations .....	48
Summary of Previous Chapters .....	48
Conclusions .....	49
Further Research .....	50
Appendix A: Raw Form R Data .....	52
Appendix B: Telephone Questionnaire .....	57
Appendix C: Raw Survey Data .....	58
Appendix D: Tally Sheets .....	60
Appendix E: Statistical Analysis (Mathcad Templates) .....	63
Bibliography .....	121
Vita .....	124

## List of Figures

Figure	Page
1. Number of Chemicals Released per Facility.....	33
2. Amount of Chemicals Released per Facility.....	34
3. Top Ten Releases to All Media.....	35
4. Distribution of Dichloromethane Releases by ALC.....	36
5. Percent breakout of Total MEK Released.....	36
6. Percent Breakout of Total Phenol Released.....	37
7. Top Ten Releases to Air.....	37
8. Top Ten Releases to Surface Water.....	38
9. Top Ten Releases to Surface Water Excluding Chlorine.....	38
10. Releases to Land.....	39
11. Number of Chemicals Released by Major Command.....	39
12. Amount of Chemicals Released per Major Command.....	40

## List of Tables

Table	Page
1. Breakdown of Releases to Air, Land and Surface Water .....	32
2. Database Contingency Table.....	41
3. SBSS vs. Non-SBSS Contingency Table.....	42
4. BEE vs. Non-BEE Contingency Table.....	42
5. Guidance Contingency Table .....	43
6. Air Force vs. Non-Air Force Guidance Contingency Table.....	43
7. Military vs. Non-Military Guidance Contingency Table.....	43
8. EPA vs. Non-EPA Guidance Contingency Table.....	43
9. Federal vs. Non-Federal Guidance Contingency Table .....	44
10. GAIA vs. Non-GAIA Guidance Contingency Table .....	44
11. Major Command Contingency Table.....	45
12. AFMC vs. Non-AFMC Contingency Table .....	45
13. Statistical Analysis of Contingency Tables.....	46



Abstract

The purpose of this research is to analyze the Air Force's 1994 Toxic Release Inventory data.

Statistical tests were used to meet the following research objectives: (1) review the CY 94 data to determine which chemicals were most often reported and which bases and major commands had the most releases in terms of the number of chemicals reported and the amount of chemicals reported and (2) investigate factors which could influence the reporting status of a facility.

An analysis of the TRI data showed that dichloromethane was the chemical released in the largest amount while methyl ethyl ketone was the most often released chemical. The five Air Logistic Centers owned by Air Force Material Command, Tinker AFB, Kelly AFB, McClellan AFB, Hill AFB and Robins AFB, were responsible for the most TRI chemicals both in terms of numbers and pounds released. An analysis of a survey of Air Force facilities indicated that two factors had an influence on the reporting status of Air Force facilities: (1) whether the TRI point of contact had confidence in the completeness of the data used for threshold computations and (2) whether the primary database used for threshold computations was the Standard Base Supply System.

# AN ANALYSIS OF 1994 AIR FORCE TOXIC RELEASE INVENTORY REPORTING

## I. Introduction

In 1986, Congress passed the Superfund Amendments and Reauthorization Act (SARA). SARA Title III is also known as the Emergency Planning and Community Right-to-Know Act (EPCRA). EPCRA requires certain industries to annually report the release of listed hazardous chemicals to the environment if certain threshold levels for manufacture, production or use are exceeded. Each facility is required to submit a separate EPA Form R for each chemical that exceeds the appropriate threshold. The United States Environmental Protection Agency (EPA) gathers, aggregates and publishes the release data each year as the Toxic Release Inventory (TRI). On 3 August 1993, President Clinton signed Executive Order (E.O.) 12856 which requires federal facilities to comply with the reporting provisions of EPCRA, among other things. All federal agencies, including the Air Force, were required to begin reporting EPCRA data to the EPA for calendar year 1994 (CY 94). Specifically, data for the TRI for CY 94 was required to be submitted by 1 July 1995 (Clinton, 1993: 41985).

Additionally, the executive order states that federal agencies should implement source reduction practices as they

develop voluntary goals to reduce the agency's total releases of toxic chemicals to the environment and off-site transfers of such toxic chemicals for treatment and disposal from facilities covered by this order by 50 percent by December 31, 1999. (Clinton, 1993: 41983).

Due to the fact that the Air Force does not have a model approach for estimating releases of toxic chemicals to the environment, generating the estimates for releases of TRI chemicals was left to each Air Force installation. The objective of this research is to analyze the Air Force TRI submission for CY 94 in order to gain an understanding of the factors that might have influenced the number and amount of chemicals reported.

#### General Issue

Since E.O. 12856 states that federal agencies set voluntary goals of 50% reduction of toxic releases by CY 99 using CY 94 as a baseline, a good understanding of the factors that influenced the number and amount of chemicals reported is necessary. In order to meet the goal, senior leadership must understand the TRI estimating process well enough to make intelligent decisions on how to reduce emissions. The data on the TRI Form R is not extensive enough to meet this requirement. Background data on the specific elements that went into the reports is necessary to better understand the entire process.

#### Problem Statement

Section 313 of EPCRA provides only general guidance on how to estimate releases of toxic chemicals to the different environmental media. Unfortunately, E.O. 12856 did not give federal facilities any more specific guidance on how to generate, collect and report the required data than did EPCRA Section 313. After E.O. 12856 was put into effect, each agency in the federal government was faced with the task of gathering and reporting TRI data by 1 July 1995. Each branch of the Department of Defense (DoD) independently set out to develop guidance on the reporting of Section 313 data. Industry

specific guidance for the TRI has been developed; however, since none of the missions and activities on Air Force bases and other federal facilities correspond one on one to the activities of the industries in the Standard Industrial Codes (SIC) identified under EPCRA, it was not possible to simply adopt the industry approaches. In February 1994, a memo from the Air Staff, "DoD Implementing Guidance for Executive Order 12856", provided initial broad guidance that base-level individuals responsible for reporting could use (USAF, 1994). During 1994, the Air Force provided EPCRA training through a contract with the GAIA Corporation to several hundred people. This training introduced all the provisions of EPCRA, including the TRI requirements, but gave little specific guidance as to what processes and chemicals should be addressed at Air Force installations. On 31 March 1995, Air Force specific TRI guidance was published by HQ USAF/CEV. The most significant contribution of this document is the determination of the exemptions that can be applied to Air Force facilities (USAF, 1995: 15-16). While this was a huge aid to the base level TRI personnel, no specific guidance has been provided to date on what TRI chemicals are typically used on Air Force installation, what common Air Force processes likely generate TRI releases and what estimating methods should be used in particular situations. Therefore, each facility has had to decide how to approach the gathering, compiling, and estimating of releases to the environment. No documents corresponding to the Title III Section 313 Release Reporting Guidance have been developed and distributed to the base level TRI personnel.

## Research Objectives

The CY 94 TRI data will be used as a baseline for the 50% reduction by CY 99 mandated by E.O. 12856. An understanding of factors that may have influenced the reporting status of Air Force facilities is needed. The data on the TRI Form R is not extensive enough to achieve this understanding. Background data on the specific elements that went into the reports is necessary to better understand the entire process. This understanding is critical to proper interpretation and utilization of reported TRI data. The results generated from this research will be valuable to the Air Staff and the EPA in developing strategies, policies and procedures to reduce emissions and meet the reduction goal. The results will also be useful to base level environmental managers as they attempt to reduce toxic chemical emissions on their facilities and share success stories between bases and major commands.

The specific objectives are as follows:

1. Review the CY 94 data to determine which chemicals were most often reported and which bases and major commands had the most releases in terms of the number of chemicals reported and the amount of chemicals reported.

- 2 Investigate factors which could influence the reporting status of a facility: Method, Training, Training Helpfulness, Start Date, Pharmacy, Primary Database, Data Completeness, Primary Guidance, List of Processes, or Process Visitation.

### Scope of Research

All federal facilities are required to report TRI information if any listed chemical breaks the appropriate threshold during the reporting period. This thesis, however, will look only at the CY 94 submissions by Air Force facilities. Additionally, it is likely that there are many factors that influence the number and amount of chemicals reported by facility. This research will only look at those factors described above. For example, the technique used to estimate releases is a factor in the amount of chemical reported. This research does not compare the different estimation techniques in terms of accuracy or cost. No interactive effects of a combination of the factors will be investigated.

This research does not address the issue of the validity or accuracy of the data reported. Comparisons of Air Force data to other organizations is not included.

## II. Literature Review

The environmental movement over the last 25 years has grown in scope and importance. A large volume of environmental laws and regulations have been promulgated during that time, many of which are compliance oriented. The Comprehensive Environmental Response Compensation and Liability Act (CERCLA or Superfund), the Resource Conservation and Recovery Act (RCRA), the Clean Water Act, and the Clean Air Act all address responsibilities of industry and government for the protection of our environmental resources. However, it is the Emergency Planning and Community Right-to-Know Act (EPCRA) that mandates the reporting of releases of toxic chemicals to the environment. The passage of EPCRA can be seen as a reaction of the U.S. Congress to the Union Carbide disaster in Bhopal, India. This incident brought the world's attention to the issue of proper chemical use and the disastrous consequences of the release of chemicals to the environment, either intentionally or unintentionally.

The data published annually in the Toxic Release Inventory (TRI) can be used for several different purposes. The original purpose of the legislation was to keep the government and public informed of what chemicals were being released into the environment (EPA, 1994a: 4). This gives the local community the ability to act as a check on the environmental stewardship of the industries in their local area. Likewise, industries can use the data as part of their environmental program to improve processes and develop new, "greener" technologies. Additionally, while there are no regulatory penalties associated with EPCRA, the TRI provides an indication of the environmental philosophy

of the industry in question. A firm's ranking on the TRI has a definite effect on the corporate environmental image that firm projects. And many firms set toxic chemical release reduction goals as part of their pollution prevention programs.

The requirement for the Air Force to generate TRI data stems directly from E.O. 12856 and EPCRA; however, the higher purpose is to inform the public of releases of toxic chemicals to the environment, so that the federal government in general, and the Air Force specifically, present the spirit of a good neighbor.

This section will describe EPCRA in detail and highlight the portions of E.O. 12856 that relate to the TRI. Next, the environmental policies and objectives of the Air Force are reviewed. Since the essence of the TRI is the estimation of releases, a wide variety of release estimating guidance will be reviewed.

### EPCRA (SARA Title III)

In 1986, Congress passed the Superfund Amendments and Reauthorization Act (SARA). SARA Title III is also known as the Emergency Planning and Community Right-to-Know Act (EPCRA) of 1986. This legislation was enacted due to growing public concern over the release of potentially dangerous chemicals to the environment and was part of the response of the American government to proactively deal with this serious issue.

EPCRA as a whole is primarily focused on:

1. identifying the amounts of chemicals present or released from facilities,
2. understanding the potential hazards the chemicals pose to the surrounding communities and the environment, and
3. providing the information to the public and local emergency planning and response organizations (GAIA, 1994: I-A3).



Section 313 of EPCRA requires industries in Standard Industrial Classification (SIC) codes 20-39 to report the release of listed hazardous chemicals to the environment on an annual basis. Under Section 313, a facility is subject to the reporting requirements if all of the following apply:

1. It has 10 or more full-time employees.
2. It conducts manufacturing operations (i.e. if it is included in SIC Codes 20 through 39.)
3. It manufactures, processes, or in any other way uses any of the listed toxic chemicals in amounts greater than the threshold quantities (EPA, 1994a: 3).

For 1989 and subsequent years, the threshold quantity for manufacturers and processors is 25,000 lbs. and for users is 10,000 lbs. This means a facility is required to report the amount of release, whether intentional or accidental, of any listed chemical that breaks the threshold (EPA, 1994a: 3).

The list of TRI reportable chemicals is dynamic; it has been modified several times since EPCRA was passed. For CY 94, there were 348 individual chemicals and 22 chemical categories on the list (EPA, 1994a: 3). In the fall of 1994, Congress added another 286 chemicals to the reportable list. This revised list goes into effect for the 1995 calendar year (EPA, 1994b: 61432).

Section 313 of EPCRA requires that releases to air, land, water and transfers to off-site facilities be reported for each toxic chemical meeting threshold reporting values. Four types of release quantity determination techniques are generally acknowledged:

1. direct measurement,
2. mass balance,
3. emission factors, and
4. engineering calculations (EPA, 1987: 2).

Direct measurement involves calculations based on measured concentrations of the chemical in a waste stream and the volume/flow rate of that stream. Mass balance is used to estimate releases around entire processes or pieces of process equipment. The amount of a chemical leaving a vessel equals the amount entering. If input and output or "product" streams are known (based on measured values), a waste stream can be calculated as the difference between input and output (any accumulation/depletion of the chemical in the equipment, e.g. by reaction, must also be accounted for). Emission factors normally express releases as a ratio of amount released to process or equipment throughput. Emissions factors, which are commonly used for air emissions, are based on the average measured emissions at several facilities in the same industry. Engineering calculations and/or judgment are based on physical/chemical properties and relationships such as the gas law (EPA, 1987: 2-6).

A trade-off exists in the selection of a particular technique. Direct measurement generates the most accurate data, but normally at a very high cost in terms of dollars and manpower. Mass balance is less accurate in estimating the true amount released, but requires less money and manpower to perform. Emission factors may suffer from an incorrect assumption that the process in question is analogous to the one from which the emission factor was developed (EPA, 1987: 6). Engineering calculations and/or judgment is the least expensive of the techniques, but has the potential for being the least accurate as well.

A general five step process which was developed to help an individual facility complete its Section 313 reports is:

1. determine if your facility processes or uses any of the chemicals subject to reporting under Section 313,
2. determine if your facility surpassed the threshold quantities established for reporting of listed chemicals last year,
3. identify points of release for the chemical(s) subject to reporting,
4. estimate releases of toxic chemicals, and
5. complete the Toxic Chemical Release Inventory Reporting Form. (EPA, 1988: 1-10).

This process can be adapted and used by all federal facilities, including the Air Force.

No specific guidance was provided in the legislation on how to develop those estimates. Therefore, industry, in cooperation with the EPA, developed chemical release estimating documents that could be applied across specific industries. The Title III Section 313 Release Reporting Guidance manuals for each industry give specific examples on how to use the techniques to measure or compute the amount released to each of the environmental media. These documents list typical chemicals involved in the process in question, show flow diagrams of the standard process, and give specific mathematical examples on how to use each of the release quantity determination techniques. Copies of these chemical release estimating documents are readily available through the EPA's EPCRA Hotline at (800)535-0202.

The mandated annual deadline for reporting under Section 313 is 1 July. After the EPA receives all the data, it is compiled by state, industry and chemical. The EPA first published the summaries of this data in the Toxic Release Inventory in 1989 and has published them annually since (Blackman, 1993:46). This is a lengthy process; for

example, the most recent TRI data available is for 1993 which was published in March 1995 (EPA, 1995b).

#### Executive Order 12856

On 3 August 1993, President Clinton signed E.O. 12856 which requires federal facilities to comply with all the planning and reporting provisions of EPCRA. Thus federal agencies are required to report TRI information in the same manner as industry. E.O. 12856 set the first reportable and baseline year as calendar year 1994 with the TRI reports due to the EPA by 1 July 1995. Additionally, E.O. 12856 requires agency heads to develop voluntary goals to reduce the agency's total release of toxic chemicals and off-site transfers by 50 percent by 31 December 1999.

#### Air Force Environmental Guidance

Even before E.O. 12856 was signed, the Air Force had taken an aggressive approach in environmental management and committed itself to be the leader in environmental protection and compliance with the DoD (McPeak, 1991). In 1993, then Secretary of the Air Force Donald Rice and Air Force Chief of Staff General Merrill McPeak issued additional policy in a joint letter. It stated:

The Air Force is committed to environmental leadership. Our goal is to prevent future pollution by reducing use of hazardous materials and releases of pollutants into the environment to as near zero as feasible. To achieve this, we must quickly move away from dependence on hazardous materials, actively reduce our waste streams, to reuse the wastes we do generate, recycle what we cannot reuse, and expand purchasing programs for recycled products. To succeed we must mobilize our whole team and find ways to move faster (Rice and McPeak, 1993).

## Release Estimating Literature

### *EPA Guidance*

In December 1987, the EPA published TRI guidance in the form of Estimating Releases and Waste Treatment Efficiencies for the Toxic Chemical Release Inventory Form. The major contribution of this document was the description of the four major approaches to estimating releases: monitoring, mass balance, emission factors and engineering estimates. The guidance then goes on to describe how to use these general estimating techniques for fugitive and point source air emissions, releases to wastewater, releases in solid, slurry, and nonaqueous liquid wastes and accidental releases (EPA, 1987).

The EPA also developed the document Guidance for Implementing Executive Order 12856: Federal Compliance with Right-to-Know Laws and Pollution Prevention Requirements. E.O. 12856 is discussed section by section along with in-depth questions and answers about the applicability of EPCRA, including Section 313, to federal facilities and the possible exemptions that might apply at any given facility (EPA, 1995b).

### *Industry Approaches*

After Congress passed SARA Title III in 1986, the effected industries immediately began to determine how to comply with the new law. As industry attempted to fulfill the requirements of EPCRA, especially Section 313, it became evident that different companies within an industry were struggling to develop release estimates from the same

or very similar processes. Therefore, with the help of the EPA, industry specific release estimation documents were developed. Examples include:

- Title III Section 313 Release Reporting Guidance - Estimating Chemical Releases From Textile Dyeing.
- Toxic Chemical Release Inventory: Clarification and Guidance for the Metal Fabrication Industry and,
- Development of Environmental Release Estimates for Welding Operations.

While these documents were not of consistent depth or quality across all the industries required to report TRI data, they all are an excellent point of departure for initial estimation determination.

#### *DoD Guidance*

The DoD published its own guidance for implementing E.O. 12856 on 13 April 1995. Each of the branches of the DoD contributed to this document. Its expressed purpose is to:

1. clarify definitions and concepts that relate directly to implementing E.O. 12856 at DoD facilities and
2. define DoD policy on implementing E.O. 12856 sections 3-302, Pollution Prevention Plans and Toxic Chemical Reduction Goals, and 3-304, Toxic Release Inventory (TRI) Reporting. (DoD, 1995).

The TRI reduction goals mandated by E.O. 12856 are addressed in depth in this document. First, the guidance states that the EPCRA TRI list of chemicals dated January 1994 will be used for DoD 50% reduction. Second, it clarifies that the baseline for reduction will be the sum of the quantity released and the quantities transferred off-site. Finally, the guidance makes clear that the 50% reduction is an agency goal and is not

facility or chemical specific. However, each facility is mandated to reduce releases and transfers as close to 50% as possible. The DoD guidance also covers the EPA approved exemptions for federal facilities, lists common questions and answers and an extensive bibliography of EPCRA and TRI resources (DoD, 1995).

### *Air Force Guidance*

EPCRA, as originally passed by Congress, was written with the industries in SIC codes 21-39 in mind. Application of the statute to non-manufacturing or defense-related facilities was not the intent of the legislation in 1986. EPCRA contains listed exemptions of certain processes that use toxic chemicals that industry did not have to include in their release calculations. For industry, this amounts to small and/or ancillary uses of listed chemicals that would not have significantly affected the amounts reported. After E.O. 12858 was signed, it became evident that under the original language contained in EPCRA, the Air Force would have been able to exempt a large majority of the processes typically found on Air Force bases. The DoD agencies reviewed the listed exemptions with the cooperation of the EPA to determine which exemptions could be applied to these facilities. Both the DoD and Air Force guidance documents list the exemptions that were approved by the EPA for federal facilities for CY 94. An understanding of the exemptions to reporting that EPA approved is critical in order to interpret the data generated by Air Force installations for calendar year 1994. Taking the exemption for a particular process has two major effects. First, the threshold level for any listed chemical involved in the exemption need not be considered. Second, if the threshold for a covered

chemical is exceeded as a result of other non-exempt activities at the facility, the facility will not be required to include the exempted amount in its report (USAF, 1995: 5).

The following categories of exemptions are summarized from the 31 March 1995 Air Force guidance and were applied by Air Force facilities for CY 94:

1. Use Exemptions

- (a) Structural Component: Toxic chemicals that are structural components of the facility or that are used to ensure or improve structural or functional integrity.

- (b) Routine Janitorial and Facility Grounds Maintenance Exemption: This exemption was developed to relieve the burden of reporting chemicals used to maintain the grounds around the plant and the day-to-day cleaning of the facility. Chemicals used to maintain the grounds of a facility's recreational areas such as golf courses are covered under this exemption.

- (c) Personal Use Exemption : Applies to the use of listed toxic chemicals in products used by employees or other persons at the facility. This exemption also includes the activities associated with facility-operated cafeterias, commissaries, exchanges, and Morale, Welfare, and Recreation (MWR).

- (d) The Motor Vehicle Exemption for Maintenance: This exemption covers toxic chemicals contained in products used for maintenance activities on motor vehicles operated by a facility. This has been interpreted by DoD to include base level vehicle maintenance (to include cars, trucks, cranes, forklifts, two motors, and boats). Depot level restoration or reconditioning is not exempt.



(e) Aircraft/Missile Maintenance: Exempted from reporting are certain aircraft maintenance activities necessary to accomplish day-to-day flying missions, such as adding hydraulic fluid, oil and lubricants, and fuel. Strict interpretation of this exemption would mean that all aircraft maintenance activities would be exempt. However, the Air Force believes that this would be inconsistent with the intent of E.O. 12856. Specifically, solvents used for parts cleaning and listed chemicals used for painting are to be counted in threshold determinations and release reporting.

(f) Motor Vehicle Maintenance Exemption Applied for Fuels: Exempted are releases associated with the transfer of fuel from stationary and non-stationary sources. In addition, emissions from mobile sources are covered under this exemption.

(g) Intake Water/Air Exemption: Any listed chemical that is in process water or no-contact cooling water drawn from the environment or from municipal sources is exempted. Additionally, toxic chemicals present in air used either as compressed air or as part of combustion are covered.

2. Laboratory Activity Exemption: Toxic chemicals manufactured, processed, or otherwise used in a laboratory under the supervision of a technically qualified individual are exempt from reporting. This exemption does not apply in the following cases:

- (1) Specialty chemical production.
- (2) Manufacture, processing, or use of toxic chemicals in pilot plant scale operations.
- (3) Activities conducted outside the laboratory.

The Laboratory Activity Exemption applies to those listed toxic chemicals manufactured, processed, or otherwise used in a laboratory for quality control, research and development, and other laboratory activities.

3. Article Exemption: Quantities of a listed toxic chemical contained in an article do not have to be factored into threshold or release determinations when that article is processed or otherwise used at a facility. An article is defined as a manufactured item that is formed to a specific shape or design during manufacture, that has an end-use function dependent in whole or in part upon its shape or design during end-use, and that does not release a toxic chemical under normal conditions of the processing or otherwise use of that item at the facility. The article exemption applies to the normal processing or otherwise use of an article. It does not apply to the manufacture of an article. Toxic chemicals processed into articles produced at a facility must be factored into threshold and release determinations.

4. De Minimis Exemption: A listed toxic chemical does not have to be considered if it is present in a mixture at a concentration below a specified de minimis level. The de minimis level is 1.0%, or 0.1% if the toxic chemical meets the Occupational Safety and Health Administration's defined carcinogen. The EPCRA Form R instructions document lists the de minimis level for all the toxic chemicals. For mixtures that contain more than one member of a listed toxic chemical category the de minimis level applies to the aggregate concentration of all such members and not to each individually.

5. Property Owners: The Air Force is not required to report if it merely owns the real estate on which a facility covered by EPCRA or E.O. 12856 is located; that is, the Air

Force has no "business" interest in the operation of that facility. The operator of that facility, however, may be subject to the reporting requirements (USAF, 1995: 2-11).

#### *GAIA Reference Materials*

During 1994, the Air Force let a contract with the GAIA Corporation to provide EPCRA training to Air Force personnel. At one time, at least two people per CONUS Air Force base had received this training (Nelson, 1994b). This does not imply however, that the TRI point of contact (POC) at any given facility received the training. The training covered all of EPCRA and the Section 313 specifics were only a part of what was presented. Therefore, it is possible that people responsible for other sections of EPCRA were the ones who received the training. Each student was provided a copy of the Emergency Planning and Community Right-to-Know Act Manual for U.S. Air Force Installations which was written by the GAIA Corporation. This guidance is notable because it provided very specific, step by step instructions on threshold determinations, estimating off-site transfers, filling out the Form R reports and what common reporting errors to avoid (GAIA, 1994). However, it did not list Air Force specific processes or attempt to indicate what chemicals might commonly be found in Air Force processes.

In early 1995, the Air Force once again contracted with GAIA Corp. for EPCRA expertise. This time the product of the contract was the Emergency Planning and Community Right-to-Know Act Handbook for U.S. Air Force Installations. This three volume set provides instructions, worksheets and examples and chemical information for

EPCRA reporting. Section 313 and the TRI are covered in depth and twelve common Air Force activities are listed as possible sources of emissions (GAIA, 1995).

### III. Methodology

The purpose of this research is to analyze the CY 94 Toxic Release Inventory data that was reported to the EPA by the Air Force. This chapter discusses the research objectives in depth, one at a time. First, the research objective is described. Second, the data collection plan for the research objective and a brief description of the data collected is provided. Finally, the statistics used to evaluate the data for the research objective are described.

#### Research Objective 1

The initial objective for this research is to review the CY 94 data to determine which chemicals were most often reported and which bases and major commands had the most releases in terms of the number of chemicals reported and the amount of chemicals reported. Percentage breakouts, pareto charts and pie charts are used to describe the following:

1. Breakdown of releases to air, land and surface water,
2. Number of chemicals released per facility,
3. Amount of chemicals released per facility,
4. Top ten releases to all media,
5. Top ten releases to air,
6. Top ten releases to surface water,
7. Top releases to land,
8. Number of chemicals released per major command, and
9. Amount of chemicals released by major command.

#### *Data Collection*

Each of the bases that reported TRI data for CY 94 were required to submit a copy of their report to the EPA. The data submitted to the EPA was also submitted to

HQ USAF/CEVV. The Air Force awarded a contract to Dynamac Corp. in Washington, D.C. to aggregate the data into a more readable version. HQ USAF/CEVV provided a copy of the aggregated data for this research. Included in the data was POC information as well as the data on the specific chemicals and amounts released at each of the reporting Air Force facilities. The Air Force TRI data was provided by HQ USAF/CEVV after Dynamac Corp. converted the electronic Form R data into Microsoft Excel spreadsheets (Appendix A). EPCRA allows facilities to update their submissions throughout the year if better estimates become available; however, no changes were made to the data as provided.

#### *Obtaining and Organizing Data*

The data provided by Dynamac Corp. through HQ USAF/CEVV was successfully received. The Form R data included the TRI POC's name and telephone number, the amount of each chemical released to air, surface water and land, and the estimation method used to develop the reported figure. The Sort function in Microsoft Excel was used to reorganize and total the data in the proper format to be analyzed. Percentage breakdowns, pareto charts and pie charts were then developed according to the research objective stated earlier.

#### *Statistical Analysis*

The data from the Form Rs was sorted according to the each of the eight criteria listed above and pareto charts were created using the sorted data. In this way, it is

possible to see which facilities, chemicals and major commands contributed significantly to the overall TRI submission.

#### Research Objective 2

Research objective 2 was to investigate factors which could influence the reporting status of a facility. This was achieved by surveying the Air Force facilities that filed Form Rs for CY 94 and a random sample of facilities that did not file for the same period. A telephone research questionnaire was used to gather this background data.

#### *Data Collection*

Prior to the reports being filed, there was no way to predict how many of the Air Force bases in the fifty U.S. states would have listed chemicals that would break the appropriate threshold and thus be required to be report. Overseas bases are not required to report TRI information. Twenty-six different facilities submitted Form Rs to the EPA and Air Staff. A single POC reported for two of the facilities, Hill AFB and the Utah Test and Training Range (UTTR), since the environmental management for both facilities are the responsibility of the same organization.

In order to complete the research, a random sample of non-reporting bases needed to be accomplished. There are 73 major Air Force facilities and 14 minor installations that come under the auspices of E.O. 12856. This does not include Air Force Reserve or Air Guard bases which were not investigated during this research. However, only one of the reporting facilities was a minor installation, Arnold Air Station. Therefore, the decision was made to limit the random sample of non-reporting bases to major installations.

Twenty-five major Air Force facilities reported TRI data which means 48 major facilities did not report any TRI releases. Each of the non-reporting facilities was numbered and the random number generator on a Hewlett Packard 15C scientific calculator was used to select a random sample of size 10, or slightly greater than 20%, of the non-reporting facilities. Twenty percent selection of non-reporting facilities was assumed to be large enough to be a representative random sample for the purpose of this research.

The background data for Air Force facilities was gathered by means of a telephone survey. Each of the 36 facilities (26 reporting and 10 non-reporting) were contacted by telephone over a two week period in September 1995. A telephone survey was performed that asked questions about the following ten factors: Method, Training, Training Helpfulness, Start Date, Pharmacy, Primary Database, Completeness of Data, Primary Guidance, List of Processes and Process Visitation. In accordance with telephone survey procedures, a written list of questions was developed that specifically asked for the categorical data of interest (Dillman, 1978: 205). The questionnaire is located in Appendix B. A short description of each of the categories follows.

1. Method: Each facility had the choice of using in-house personnel to gather and report TRI or contracting the operation out. The number or amount of chemicals reported could have been affected by the decision of whether to perform the work in-house or by contract.

2. Training: During 1994, the GAIA Corporation provided EPCRA training to the Air Force. This category deals with whether the TRI POC attended this training or



not. Whether the TRI POC had this formal training could have affected the reporting status of the facility.

3. Training Helpfulness: For those TRI POCs who did receive the GAIA training, this factor queries how helpful that training was to the CY 94 TRI process. The perceived usefulness of the training to the TRI POC may have had an influence on the number or amount of chemicals reported.

4. Start Date: This factor deals with when the facility started its TRI reporting process. The time allotted to determining which chemicals broke the threshold, calculating release estimates and filling out the Form Rs could have had an effect on the reporting status of the facility.

5. Pharmacy: Did the facility have a fully functional Hazardous Material Pharmacy in operation for all of CY 94? A Hazardous Material Pharmacy is the single focal point for the supply and issue of all hazardous materials on an Air Force base, from issue to final disposal (Nelson, 1994a: 3). The presence of a fully functional Hazardous Material Pharmacy could have had an effect on the reporting status of the facility.

6. Primary Database: This factor is concerned with the source of data for determining whether the threshold for any particular chemical was exceeded. The possibilities for major data sources are the Standard Base Supply System, which is the main source of supplies on all Air Force facilities, databases maintained by Bio-Environmental Engineering (BEE), which is a branch of the base hospital that documents the presence of hazardous materials on base, Hazardous Material Pharmacy databases, which are designed to track hazardous materials from the time they come on base to their

ultimate disposal and the Civil Engineering Material Acquisition System, which the Civil Engineer squadron on base uses as its main vehicle for the acquisition of materials and supplies. Each of these databases contains different data. The choice of any single database could have an effect on the number and amount of chemicals reported.

7. Data Completeness: The TRI POC was asked if they felt the data from the primary database appeared to be complete or not. If the decision was made that the data from the chosen primary database was sufficient, it may have been possible that the reporting status of the facility could have been affected.

8. Primary Guidance: The intent behind this factor is to determine the primary source of guidance for developing, aggregating and reporting the TRI data. Possible sources include Air Force guidance, DoD guidance, E.O. 12856, two different guidance documents published by the GAIA Corp., and the electronic Form R instructions. While all of these documents discuss EPCRA and the TRI, they do not all contain the same material or emphasis the same issues. Since the primary guidance document is the most likely source for information on how to make estimates or what chemicals and processes to consider, the choice of which document to use may have an effect on the reporting status of the facility.

9. List of Processes: This factor deals with whether the TRI POC developed a list of processes on the facility that generate TRI releases. By developing a list of the TRI chemical generating processes on the base, the TRI POC is documenting a certain level of investigative effort into the release of chemicals to the environment. This level of effort may be reflected in the reporting status of the facility.

10. Process Visitation: The thrust here is to determine if the TRI POC visited the process sites where TRI releases take place. Again, this is a reflection of level of effort. Visiting the process sites shows at least a minimum investment of time and effort. The results of this investment may manifest itself in the number or amount of TRI chemicals reported.

Each question was designed so that the answer would fall into only one category. For example, the first question asked if the facility's TRI process was performed by in-house personnel or if a contract was awarded to develop the data. The answers fell into one of the two categories: In-house or Contract. A similar process was used for the rest of the questions on the survey. The summary matrix of the answers is located in Appendix C.

A tally sheet for each question was developed. The answers given by the TRI POC were recorded in the proper row and column on the appropriate tally sheet. The entries in each row were totaled and checked to ensure that the row for the facilities that filed Form Rs equaled 26 and row total for those bases that did not report any TRI releases equaled 10. This check was done to help ensure there were no data points missing. These tally sheets could then be used as a contingency tables.

#### *Obtaining and Organizing Data*

The TRI POC at each of the 26 facilities was contacted by telephone. In several cases, the POC identified on the Form R was no longer working in the that capacity. In those cases, another person who was familiar with the facilities TRI reporting process for

CY 94 answered the survey questions. For the 10 non-reporting bases, the secretary of Environmental Management office was asked to provide the name of the TRI POC for their base. Those people identified were then contacted and confirmed as the appropriate person to answer the survey questions. All 36 facilities answered all the questions on the survey.

A tally sheet for each question was used. The answers the TRI POC gave were recorded in the appropriate block on the tally sheet depending on what answer the TRI POC gave to the question and whether the facility had filed a TRI report. After all 36 facilities had answered the survey questions, the rows were totaled to check for accuracy. For this research, the row total for the bases that reported TRI releases is fixed at 26 and row total for the bases that did not report is fixed at 10. Once the row totals had been verified, these tally sheets were in contingency tables format and could be analyzed using contingency table analysis (Appendix D).

The numbers from the tally sheets were then entered into the Mathcad template. The template automatically calculated the expected values and compared the P-value against the  $\alpha$  value of 0.10. The Mathcad templates for each of the questions can be found in Appendix E.

### *Statistical Analysis*

Contingency tables are data tables which are created whenever categorical data are cross-classified. If each member of a sample is classified by one characteristic into R classes and by a second characteristic into C classes, the data may be presented in a table

with R rows and C columns. The entry in any of the RC cells is the number of members of the sample falling into that cell (Snedecor and Cochran, 1980: 208-209).

Categorical data are generally thought to consist of two distinct types of measurement:

1. nominal measurements: simple counts, labels and names (for example: male, female, British, American) and
2. ordinal measurements: counts, labels and names which exhibit a qualitative relationship or order rank (O'Brien, 1989: 18).

When cross-classifying, three distinct types of contingency tables may be created:

1. fully nominal tables created by the cross-classification of two or more nominal variables,
2. mixed contingency tables created by the cross-classification of nominal and ordinal variables, and
3. fully ordinal tables created by the cross-classification of two or more ordinal variables (O'Brien, 1989: 19).

Most often, the total sample is assumed to have been drawn before the entities were classified according to the criteria of classification. That is, the observed number of entities falling into each cell was determined after the sample was drawn. As a result, the row and column totals are random variables not under the control of the investigator. The sample drawn under these conditions is considered a single sample drawn for a single population. This leads to a chi-square test of independence (Daniel, 1991: 551).

On occasion, however, either row or column totals may be under the control of the investigator; that is, the investigator may specify that independent samples be drawn for each of several populations. In this case one set of marginal totals is said to be fixed, while the other set, corresponding to the criterion of classification applied to the samples is random. This situation leads to a chi-square test of homogeneity (Daniel, 1991: 552).

The two situations not only involve different sampling procedures; they lead to different questions and null hypotheses. The test of independence is concerned with the question: Are the two criteria of classification independent? The homogeneity test is concerned with the question: Are the samples drawn from populations that are homogeneous with respect to some criterion of classification? In the latter case, the null hypothesis states that the samples are drawn from the same population. Despite these differences in concept and sampling procedure, the two tests are mathematically identical (Daniel, 1991: 552).

In the case of testing for independence, the total sample was assumed to have been drawn before the entities were classified according to the criteria of classification. That is, the observed number of entities falling into each cell was determined after the sample was drawn. As a result, the row and column totals are chance quantities not under the control of the investigator. The characteristics of a chi-square test of independence that distinguish it from other chi-square tests are as follows:

1. Generally, a single sample is selected from a population of interest and the subjects or objects are cross-classified on the basis of the two variables of interest.
2. The rationale for calculating the expected cell frequencies is based on the probability law which states that if two events (here the two criteria of classification ) are independent, the probability of their joint occurrence is equal to the product of their individual probabilities.
3. The hypotheses and conclusion are states in terms of the independence (or lack of independence) of two variables (Daniel, 1991: 548-549).

The second case concerns test of homogeneity. The statistical use of the chi-square is exactly the same as in the test for independence; however, different assumptions and interpretations are in order.

The sampling technique used for tests of homogeneity are called product-multinomial. For each category of the row variable, a multinomial is taken and each member of the sample is classified according to its category for the column variable (Fienberg, 1978: 15). In this manner, the row totals are fixed marginal totals and the column totals are random variables.

The chi-square test for homogeneity that is used with contingency tables with fixed marginal totals has the following assumptions associated with it:

1. Each observation is classified into exactly one cell.
2. The observations are observations on a random sample. Each observation has the same probability of being classified into cell  $(r,c)$  as any other observation.
3. The row or column totals are given, not random (Conover, 1980: 162).

The chi-square test of homogeneity has the following characteristics:

1. Two or more populations are identified in advance and an independent sample is drawn from each.
2. Sample subjects or objects are placed in appropriate categories of the variable of interest.
3. The calculation of expected cell frequencies is based on the rationale that if the populations are homogeneous as stated in the null hypothesis, the best estimate of the probability that a subject or object will fall into a particular category of the variable of interest can be obtained by pooling the sample data.
4. The hypotheses and conclusions are stated in terms of homogeneity (with respect to the variable of interest) of populations (Daniel, 1991: 553-554).

Unfortunately, the chi-square test becomes unreliable when numerous expected cell values are near zero. This leads to the following general rules:

1. No expected values should be less than one.

2. Two expected values may be close to one if most of the other expected values exceed five.
3. Classes with expectations less than one should be combined to meet the first two rules (Snedecor and Cochran: 1980, 77).

When performing statistical analysis, a decision must be made on the preferred level of significance for the tests performed. For this research, the significance level,  $\alpha$ , for all statistical tests was set at 0.10. The significance level is the probability of a Type I error which is defined as rejecting the null hypothesis,  $H_0$ , when it is true (Devore, 1991: 286). Devore states that an  $\alpha$  value of 0.10 is a traditional level of significance - although a smaller significance level should be used if a Type I error can be considered very serious. For this research, an  $\alpha$  of 0.10 was selected.

The  $P$ -value is the smallest level of significance at which  $H_0$  would be rejected when a specified test procedure is used on a five data set. Once the  $P$ -value has been determined, the conclusion results from comparing the  $P$ -value to  $\alpha$ :

1.  $P\text{-value} \leq \alpha \Rightarrow \text{reject } H_0 \text{ at level } \alpha$
2.  $P\text{-value} > \alpha \Rightarrow \text{do not reject } H_0 \text{ at level } \alpha$  (Devore, 1991: 315).

A Mathcad 5.0 Plus template was created that covers all the statistical steps of analyzing the contingency tables. The template was altered to reflect each question and the results of the statistical analysis.



#### IV. Findings and Analysis

This chapter reports the results of the statistical analysis performed to gain an understanding of the factors that might have influenced the number and amount of TRI chemicals reported by Air Force facilities for CY 94. The results are explained in two sections. First, the results of the statistical analysis is presented for each research objective and second, an evaluation of the overall analysis is provided.

##### Results of Statistical Analysis

###### *Research Objective 1*

The initial objective for this research is to review the CY 94 data to determine which chemicals were most often reported and which bases and major commands had the most releases in terms of the number of chemicals reported and the amount of chemicals reported. Tables, pareto charts and pie charts were developed for the different areas of interest. The first is a table that shows the distribution of the total Air Force releases of TRI chemicals in pounds that were released to the different media (Table 1). An overwhelming percentage of the releases were to air: 98.9%. Only 1.73% of the releases were to land and 0.37% to surface water.

Table 1. Breakdown of Releases to Air, Land and Surface Water

Media	Pounds Released	Percent of Total
Air	3,710,677	97.90%
Land	65,677	1.73%
Surface Water	13,923	0.37%
<b>Total</b>	<b>3,790,277</b>	

The extremely large percentage of releases might be an indication of where the emphasis was placed for reporting TRI data. However, none of the TRI documentation emphasize the reporting of releases to the air more than releases to the other environmental media. Other possibilities exist for the skew in the data. It is possible Air Force policy has been to emphasize reducing water and land emissions under other environmental programs.

Next is a pareto chart that displays the number of chemicals released per facility. Figure 1 shows this result. Tinker AFB reported the most chemicals, 23, followed by Hill AFB with 17, Kelly AFB with 10, McClellan AFB with nine and Edwards AFB and Robins AFB with six each. All of these bases are Air Force Material Command (AFMC) bases and, with the exception of Edwards AFB, all are Air Logistics Centers (ALCs).

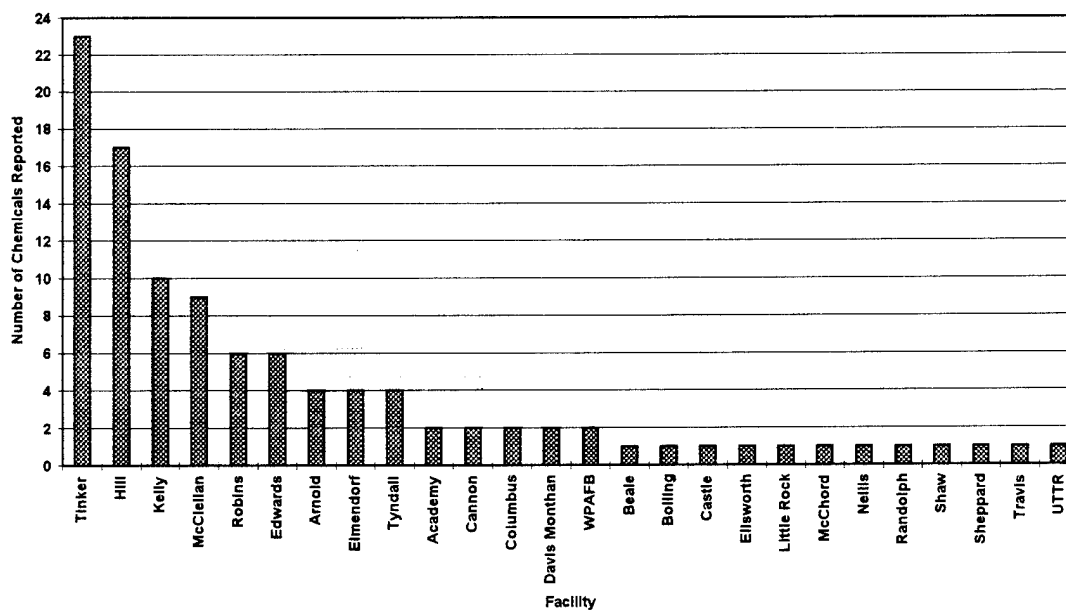


Figure 1. Number of Chemicals Released per Facility

The amount of chemicals released by each facility is shown in Figure 2. Tinker AFB released the largest amount of chemicals in terms of pounds. It was followed by its fellow AFMC facilities: Robins AFB, Hill AFB, McClellan AFB and Kelly AFB. As with the previous chart, the ALCs led the way.

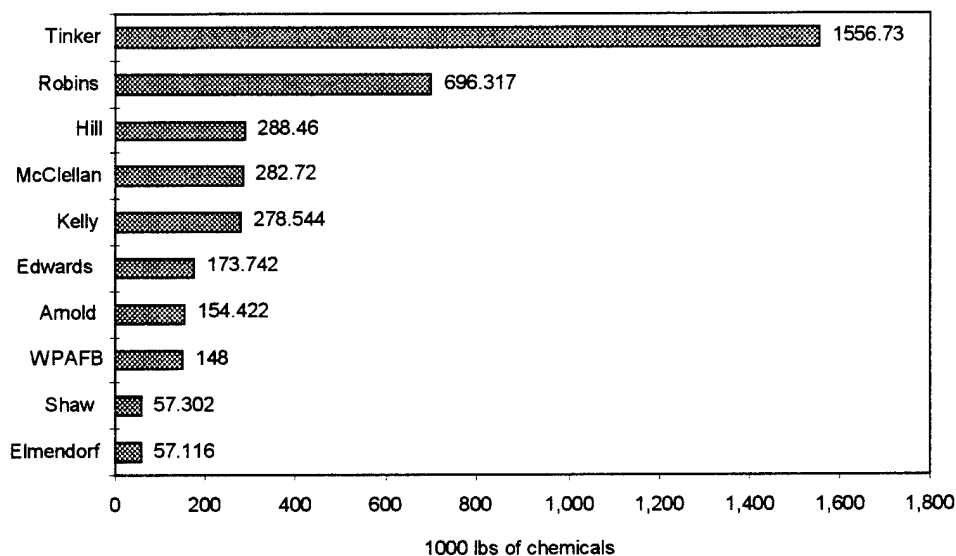


Figure 2. Amount of Chemicals Released per Facility

What is it about the ALCs that place them at the top of these lists? . ALCs provide life-cycle weapon system sustainment, maintenance, and repair for the Air Force major weapon systems (Air Force Magazine, 1995: 79). This is commonly known as depot-level maintenance versus the base-level maintenance that takes place day-to-day in active flying wings. The ALCs are responsible for complete overhauls of the weapon systems, therefore it is likely they use more materials and supplies than a typical flying wing, including TRI chemicals. Also, there are some maintenance activities that only

occur at the depot level, therefore, the ALCs might be the only Air Force facilities that use certain TRI chemicals.

The top ten releases to all environmental media are shown in Figure 3. Dichloromethane was the top chemical in terms of amount released to all the environmental media with 1,498,301 pounds. MEK was second with 509,360 pounds released. Phenol was next at 340,308 pounds.

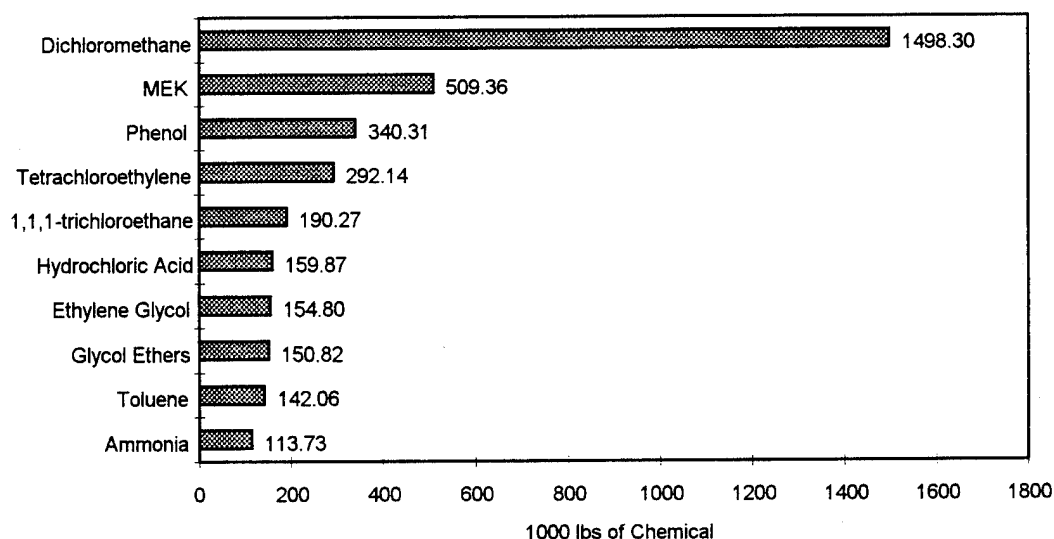


Figure 3. Top Ten Releases to All Media

The ALCs were responsible for 100% of the release of dichloromethane to the environment. Figure 4 shows the break out for each facility. This seems to indicate that an ALC-specific process exists which uses dichloromethane that is not found on non-ALC bases.

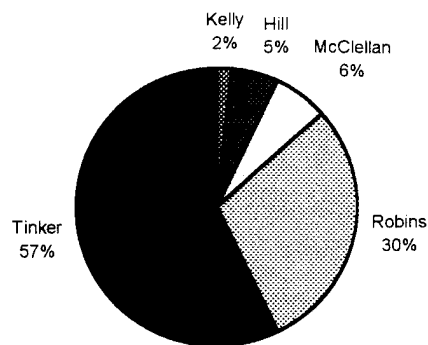


Figure 4. Distribution of Dichloromethane Releases by ALC

Methyl ethyl ketone not only was second in terms of pounds released at 509,360 pounds, but first in terms of being the most often reported chemical with 15 of the 26 facilities reporting a release of MEK to the environment. All five of the ALCs reported MEK releases and those releases accounted for 74%, or 379,265 pounds, of the total MEK released (Figure 5). Ten other facilities combined to release 130,095 pounds of MEK.

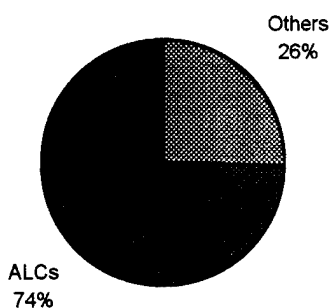


Figure 5. Percent breakout of Total MEK Released

Phenol had the third largest total release to the environment with 340,308 pounds released. The five ALCs accounted for 100% of the release of phenol to the environment. Figure 6 shows the percentage breakout for those five facilities.

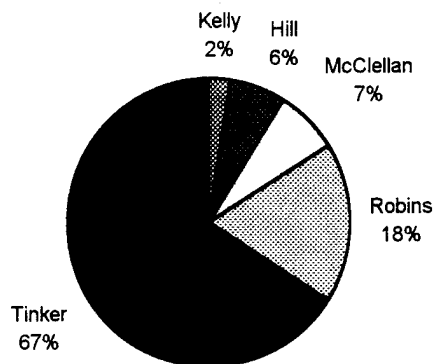


Figure 6. Percent Breakout of Total Phenol Released

Dichloromethane was also the chemical with the largest release to air. Figure 7 shows the top ten releases to air. MEK was second and phenol third.

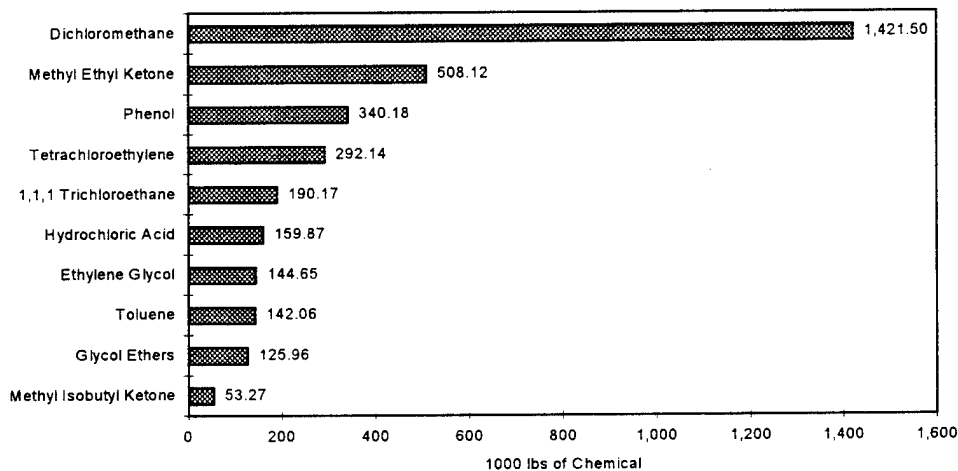


Figure 7. Top Ten Releases to Air

The top ten releases to surface water are shown in Figure 8. Chlorine was far and away the chemical with the largest amount of released with 12,200 pounds. Beale AFB was responsible for 9,000 of that amount with the other 3,200 coming from Ellsworth AFB. Ethylene glycol was second with 750 pounds, all from Hill AFB and sulfuric acid was third at 326 pounds all from Arnold AS.

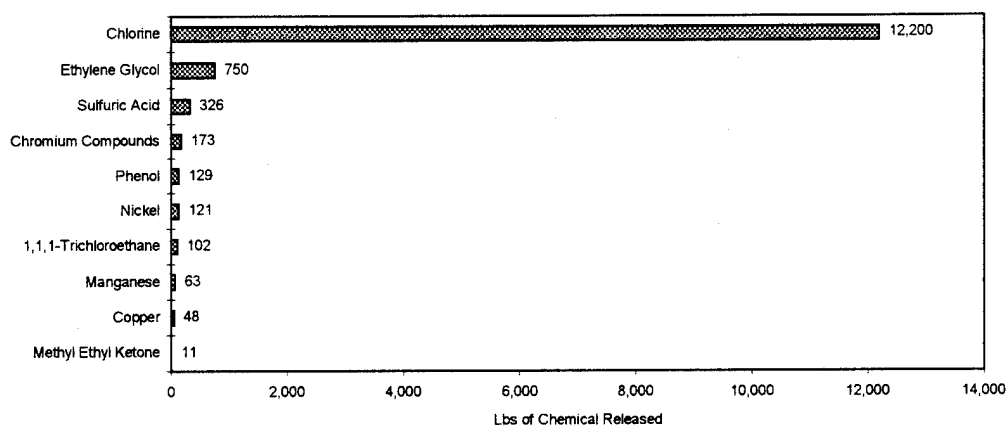


Figure 8. Top Ten Releases to Surface Water

Figure 9 shows the releases to surface water without the chlorine data. This shows that ethylene glycol was a much larger release than the other chemicals.

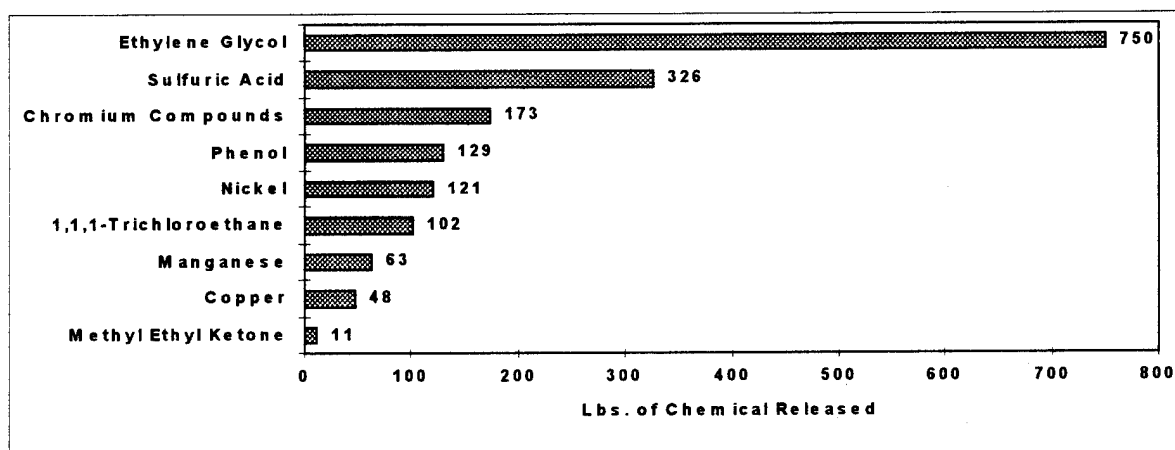


Figure 9. Top Ten Releases to Surface Water Excluding Chlorine

The largest release to land was ammonia from Elmendorf AFB at 45,000 pounds. Glycol ethers were second, followed by ethylene glycol, chlorine and MEK. This data is shown in Figure 10.

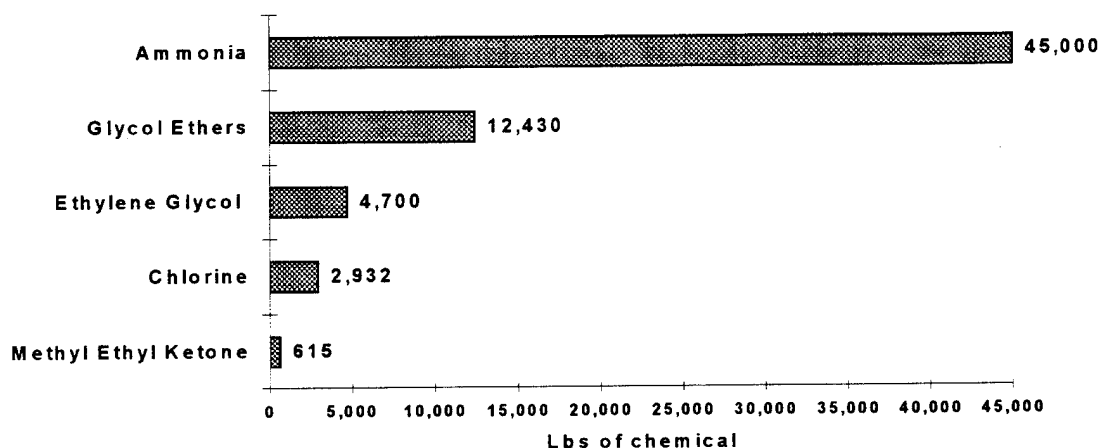


Figure 10. Releases to Land

Figure 11 shows the number of chemicals released per major command. As could be expected from Figure 1, AFMC bases lead the way with 66 of TRI chemicals. Both AMC and ACC released 10 different TRI chemicals each.

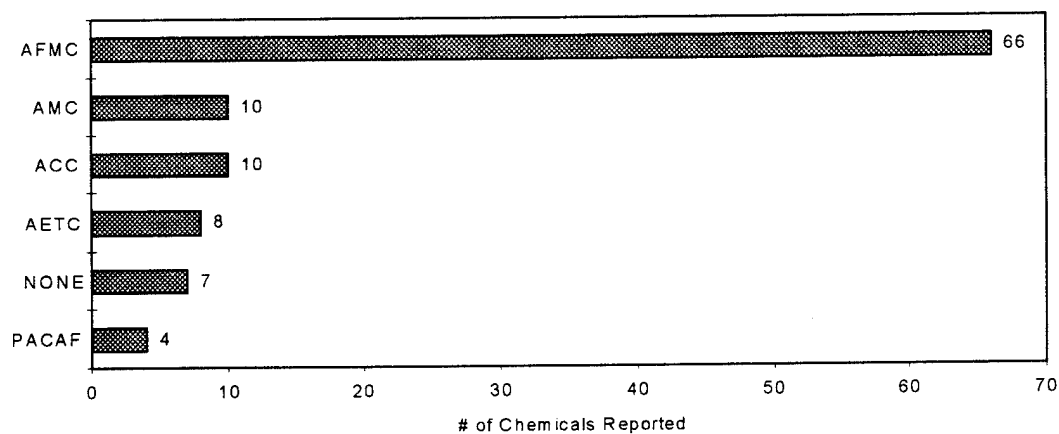


Figure 11. Number of Chemicals Released by Major Command



In terms of pounds of TRI chemicals released, AFMC again led the way with 3,424,510 pounds. ACC was a distance second with 171,790. The facilities that do not report to any major command, the Air Force Academy, Arnold AFB and Bolling AFB, were next with a total of 159,670 pounds. This is shown in Figure 12.

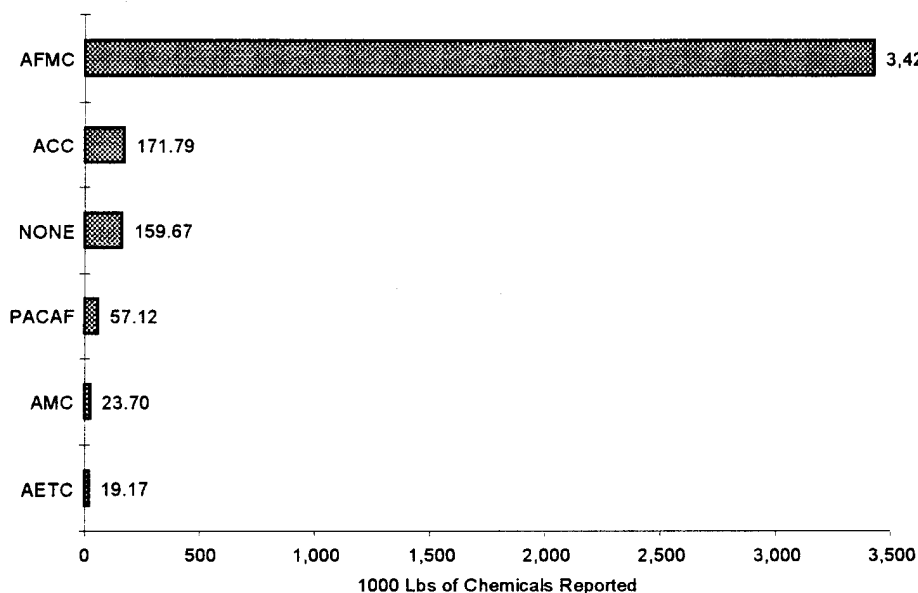


Figure 12. Amount of Chemicals Released per Major Command

### *Research Objective 2*

Research Objective 2 was to investigate factors which could influence the reporting status of a facility. The factors considered were Method, Training, Training Helpfulness, Start Date, Pharmacy, Primary Database, Data Completeness, Primary Guidance, List of Processes, and Process Visitation.

The statistical analysis of the contingency tables created by the tally sheets for these factors showed that several of the contingency tables violated the rules given earlier:

1. No expected values should be less than one.
2. Two expected values may be close to one if most of the other expected values exceed five.
3. Classes with expectations less than one should be combined to meet the first two rules. (Snedecor and Cochran: 1980, 77).

The contingency tables that violated these rules were for Primary Database and Guidance. Therefore, these contingency tables were collapsed and several new contingency tables were created. Under the Database category, two two-by-two tables were created: SBSS vs. Non-SBSS and BEE vs. Non-BEE. The original Database contingency table is shown as Table 2 and the new tables are shown as Table 3 and Table 4.

Table 2. Database Contingency Table

<b>Reported?</b>	<b>Database</b>				
	<b>CEMAS</b>	<b>SBSS</b>	<b>Pharmacy</b>	<b>BEE</b>	<b>Other</b>
<b>Yes</b>	1	7	3	13	2
<b>No</b>	0	6	0	4	0

Table 3. SBSS vs. Non-SBSS Contingency Table

<u>Reported?</u>	<u>Database</u>	
	SBSS	Non-SBSS
Yes	7	19
No	6	4

Table 4. BEE vs. Non-BEE Contingency Table

<u>Reported?</u>	<u>Database</u>	
	BEE	Non-BEE
Yes	13	13
No	4	6

The first of these new contingency tables were designed to test if there was an influence on the reporting status of the facility if the Standard Base Supply System database was used to determine if threshold levels were exceeded. The second does the same in the case of the database maintained by the Bio-Environmental Engineering function of the base hospital. The decision to collapse the original table into these two tables was based on the fact that both reporting and non-reporting facilities had indicated they had used these two databases as their primary source of data.

Under the Guidance category, four two-by-two tables were created: Air Force vs. Non-Air Force, Military vs. Non-Military, EPA vs. Non-EPA, Federal vs. Non-Federal and GAIA vs. Non-GAIA. The original Guidance contingency table is shown as Table 5 and the new contingency tables are shown as Tables 6, 7, 8, 9, and 10, respectively.

Table 5. Guidance Contingency Table

Reported?	Guidance						
	AF	DoD	GAIA Manual	GAIA Handbook	EPA	State	Form R
Yes	11	1	4	1	2	1	6
No	5	0	1	2	0	1	1

Table 6. Air Force vs. Non-Air Force Guidance Contingency Table

<b>Reported?</b>	<b>Guidance</b>	
	AF	Non-AF
Yes	11	15
No	5	5

Table 7. Military vs. Non-Military Guidance Contingency Table

<b>Reported?</b>	<b>Guidance</b>	
	Military	Non-Military
Yes	15	14
No	5	5

Table 8. EPA vs. Non-EPA Guidance Contingency Table

<b>Reported?</b>	<b>Guidance</b>	
	EPA	Non-EPA
Yes	8	18
No	1	9

Table 9. Federal vs. Non-Federal Guidance Contingency Table

<b>Reported?</b>	<b>Guidance</b>	
	Federal	Non-Federal
Yes	20	6
No	6	4

Table 10. GAIA vs. Non-GAIA Guidance Contingency Table

<b>Reported?</b>	<b>Guidance</b>	
	GAIA	Non-GAIA
Yes	5	21
No	3	7

For example, the first of these tables were designed to test if there was an influence on reporting status if the guidance used was produced by the Air Force or not. The second was created by combining the responses for the Air Force and the DoD guidance documents and labeling them as Military. The EPA vs. Non-EPA contingency table was produced by combining the EPA and Form R columns. The columns associated with the two GAIA guidance documents were combined in creating the GAIA vs. Non-GAIA table and the last new table was created when the Air Force, DoD, EPA and Form R responses were combined to form the Federal vs. Non-Federal table.

Because of the results of research objective 1 which showed that AFMC bases, specifically the ALCs, reported a significant portion of the TRI releases, contingency

tables were created to test whether the Major Command the facility belongs to had an effect on their reporting status. The first contingency table in this group compares all the different represented commands. It is shown as Table 11. The second, a two-by-two table, compares whether there was an influence on the reporting status if the facility belongs to Air Force Material Command (AFMC) or not (Table 12).

Table 11. Major Command Contingency Table

<b>Reported?</b>	<b>Major Command</b>					
	ACC	AFMC	AMC	PACAF	AETC	None
Yes	8	8	2	1	4	3
No	3	2	3	0	2	0

Table 12. AFMC vs. Non-AFMC Contingency Table

<b>Reported?</b>	<b>Major Command</b>	
	AFMC	Non-AFMC
Yes	8	18
No	2	8

The Mathcad templates for the statistical analysis of all the contingency tables are included in Appendix E.

Table 13 shows the results of the statistical analysis of the contingency tables created from the survey data. It includes information from all the categories discussed thus far.

Table 13. Statistical Analysis of Contingency Tables

Question Area	P-value	$\alpha$	Accept or Reject Ho
Method	0.9	0.1	Accept
Training	0.763	0.1	Accept
Training Helpfulness	0.735	0.1	Accept
Start Date	0.603	0.1	Accept
Pharmacy	0.379	0.1	Accept
Primary Database	0.199	0.1	<i>Violates Rules</i>
SBSS	<b>0.064</b>	0.1	<b>Reject</b>
BEE	0.59	0.1	Accept
Data Completeness	<b>0.047</b>	0.1	<b>Reject</b>
Primary Guidance	0.571	0.1	<i>Violates Rules</i>
AF	0.677	0.1	Accept
Military	0.836	0.1	Accept
EPA	0.197	0.1	Accept
Federal	0.31	0.1	Accept
GAIA	0.486	0.1	Accept
List of Processes	0.529	0.1	Accept
Process Visitation	0.958	0.1	Accept
MAJCOM	0.477	0.1	Accept
AFMC	0.518	0.1	Accept

As can be seen from Table 13, only two of the P-values generated by the statistical analysis indicated that the null hypothesis should be rejected. The first is the contingency table associated with the question of whether the TRI POC's assessment of whether the data from the primary database was complete or not had an influence on the reporting status of the facility. From the statistical analysis, it appears that the TRI POC's assessment did have an influence on the reporting status. The second is the contingency table associated with whether the primary database used for the threshold determination was the SBSS or some other database. From the statistical analysis, it appears that if the SBSS was primary database used, it had an influence on the reporting status of the facility.

Unexpected, however, are the results from the MAJCOM and AFMC vs. Non-AFMC contingency tables. The statistical analysis of these contingency tables does not

indicate that the Major Command the facility belongs to had an effect on its reporting status. The same can be said for whether the base was an AFMC facility. From the results of the analysis of research objective 1, it would have appeared that these factors would have had an influence.



## V. Conclusions and Recommendations

This chapter presents the conclusions and recommendations of this study. First, Chapters 1 through 3 and the findings from Chapter 4 are summarized. Second, the conclusions drawn from these findings are discussed. Finally, some recommendations for follow-on research are highlighted.

### Summary of Previous Chapters

The purpose of this research is to analyze the 1994 Toxic Release Inventory data that was reported to the EPA by the Air Force. Chapter 1 presented a brief introduction to EPCRA and E.O. 12856; discussed the 50% reduction goal; highlighted the lack of specific estimating guidance; and stated the research objectives of this effort. The two research objectives are: (1) review the CY 94 data to determine which chemicals were most often reported and which bases and major commands had the most releases in terms of the number of chemicals reported and the amount of chemicals reported and (2) investigate factors which could influence the reporting status of a facility which are: Method, Training, Training Helpfulness, Start Date, Pharmacy, Primary Database, Data Completeness, Primary Guidance Used, List of Processes, or Process Visits. Chapter 2 is a review of the literature that presented the legislative history of the TRI, Air Force environmental guidance and a thorough discussion of the release estimating literature that is available. The methodology described in Chapter 3 outlines the methods used to collect and test the data to meet the research objectives.

Chapter 4 presented the findings of the statistical analysis of the data for each research objective. For research objective 1, dichloromethane, methyl ethyl ketone, chlorine and ammonia stood out as chemicals of primary concern. The Air Logistics Centers, Tinker AFB, Robins AFB, Kelly AFB, McClellan and Hill AFB, were found to be responsible for the overwhelming majority of TRI releases both in terms of number and pounds released. Each of the 10 factor categories for research objective 2 were statistically analyzed using the chi-squared test for homogeneity. The statistical analysis showed that the completeness of the data and the use of the SBSS database for threshold determinations had an influence on the reporting status of the facility.

### Conclusions

In developing the strategy to reach the goal of 50% reduction of toxic chemical releases by CY 99, the Air Force should focus its efforts on air emissions, especially dichloromethane, methyl ethyl ketone, chlorine and ammonia as far as chemicals are concerned. Since dichloromethane contributes largely to the air emissions, a proportionate amount of the reduction effort should be focused on reducing these releases. However, since the Air Logistics Centers, Tinker AFB, Robins AFB, Kelly AFB, McClellan and Hill AFB, are responsible for the overwhelming majority of releases both in terms of number and pounds released, the primary focus of any reduction strategy must be those five facilities. Although the statistical analysis showed that the completeness of the data and the use of the SBSS database for threshold determinations had an influence on the reporting status of the facility, the basis for this finding is a statistical level of

significance equal to 0.10. If a larger level of significance is used, 0.20 for example, then other factors appear to have an influence. In the case of  $\alpha = 0.20$ , the overall Database category and whether the guidance was EPA generated both would appear to influence reporting status. The results of the tests for homogeneity for the other eight categories seem to indicate that the reporting status of Air Force facilities was not affected by those factors. Caution should be used however, since CY 94 was the first year of reporting and a typical learning curve effect is likely to be seen in the CY 95 data.

#### Further Research

More research is needed in this area. Since the results of this effort are being published in late CY 95, the TRI process for the next year's TRI submission is well underway. A comparative study between CY 94 and CY 95 data would indicate if learning and process improvement had taken place between reporting years. Development of a standard process-specific guidance document would contribute greatly to ensuring TRI submissions from Air Force facilities are consistent and thus comparable and measurable. This effort only concentrated on Air Force TRI submissions. Since the reduction goal is aimed at the DoD level, an effort similar to this one DoD-wide would also be useful.

Additionally, some other issues need to be considered. What processes at the ALCs produce emissions of dichloromethane, MEK, and phenol in such large volumes? Do other base have these processes or are they only ALC-specific? If these processes exist at other Air Force facilities, why didn't those facilities report releases of these chemicals in a similar proportion to the ALCs?

Still more questions need to be addressed. Concerning air releases, why was the percentage of emissions to air so high (97.9%). Has the Air Force concentrated on managing water and land emissions? Concerning surface water releases, why did only Beale AFB and Ellsworth AFB report chlorine releases to surface water. Chlorine disinfection of drinking water is a very common process. Why didn't other bases report chlorine releases?

Appendix A: Raw Form R Data

CY 94 TRI Emissions to Air						
Installation	Chemical/Chemical Category	Fugitive Emissions (lbs)	BOE Fug	Stack Emission (lbs)	BOE Stack	Total
Arnold	Ethylene Glycol	120,996	C			120,996
Arnold	Dichlorodifluoromethane	33,100	C			33,100
Arnold	Chlorine					
Arnold	Sulfuric Acid					
Edwards	Formaldehyde	245	O	51,200	E	51,445
Edwards	Hydrochloric Acid	245	O	38,116	E	38,361
Edwards	Toluene	6,898	O	27,758	E	34,656
Edwards	Propylene	5	O	32,043	E	32,048
Edwards	Glycol Ethers	245	O	1,380	O	1,625
Edwards	Methyl Tert-Butyl Ether	245	O			245
Hill	1,1,1 Trichloroethane	85,000	O	13,000	O	98,000
Hill	Dichloromethane	5,800	E	71,000	E	76,800
Hill	Methyl Ethyl Ketone	6,000	E	18,000	E	24,000
Hill	Phenol	1,100	O	21,000	O	22,100
Hill	Methyl Isobutyl Ketone	750	E	15,000	E	15,750
Hill	Toluene	7,000	E	6,200	E	13,200
Hill	Xylene (mixed isomers)	4,500	E	2,400	E	6,900
Hill	Sulfuric Acid	6,000	O	245	O	6,245
Hill	Ethylene Glycol	5,200	O	245	E	5,445
Hill	1,2,4-Trimethylbenzene	2,300	E	245	E	2,545
Hill	Glycol Ethers	1,100	O	1,100	O	2,200
Hill	Cyclohexane	1,400	E	245	E	1,645
Hill	Benzene	1,000	E	245	E	1,245
Hill	Ethylbenzene	1,000	E	245	E	1,245
Hill	Napthalene	245	E	245	E	490
Hill	Chromium Compounds	5	E	245	E	250
Hill	Lead	245	O	5	E	250
Kelly	Tetrachloroethylene	182,226	O	22,523	O	204,749
Kelly	Methyl Ethyl Ketone			30,367	O	30,367
Kelly	Dichloromethane	6,551	O	19,652	O	26,203
Kelly	Phenol	1,991	O	5,973	O	7,964
Kelly	Toluene			6,752	O	6,752
Kelly	Hydrochloric Acid			1,512	O	1,512
Kelly	Nickel Compounds	758	O	14	O	772
Kelly	Chromium Compounds	80	O	60	O	140
Kelly	Sulfuric Acid					0
Kelly	Nitric Acid					0
McClellan	Dichloromethane	95,000	C	1,000	C	96,000
McClellan	Methyl Ethyl Ketone	20,000	C	60,000	C	80,000
McClellan	Glycol Ethers	21,000	C	13,000	C	34,000
McClellan	Phenol	25,000	C	250	C	25,250
McClellan	Toluene	9,900	C	8,600	C	18,500
McClellan	Methyl Isobutyl Ketone	7,200	C	5,100	C	12,300
McClellan	Tetrachloroethylene	12,000	C	120	C	12,120
McClellan	1,1,1-Trichloroethane	2,500	C	2,000	C	4,500
McClellan	Zinc Compounds	50	E			50
Robins	Dichloromethane	436,202	C	6,546	C	442,748
Robins	1,1,1-Trichloroethane	78,497	C			78,497
Robins	Phenol	60,736	C	579	C	61,315
Robins	Methyl Ethyl Ketone	34,452	C	24,678	C	59,130
Robins	Toluene	22,501	C	11,371	C	33,872
Robins	Methyl Isobutyl Ketone	18,281	C	2,329	C	20,610

CY 94 TRI Emissions to Air (Cont.)						
Installation	Chemical/Chemical Category	Fugitive Emissions (lbs)	BOE Fug	Stack Emission (lbs)	BOE Stack	Total
Tinker	Dichloromethane	1,180	O	855,370	O	856,550
Tinker	Phenol	170	O	223,380	O	223,550
Tinker	Methyl Ethyl Ketone	21,424	O	164,333	O	185,757
Tinker	Glycol Ethers	81,970	O	6,160	O	88,130
Tinker	Tetrachloroethylene	5,664	O	69,608	O	75,272
Tinker	Ammonia	23,724	O	4	O	23,728
Tinker	Ethylene Glycol	18,208	O			18,208
Tinker	Toluene	7,186	O	10,632	O	17,818
Tinker	Xylene (mixed isomers)	12,668	O	4,222	O	16,890
Tinker	Freon 113	15,021	O			15,021
Tinker	1,1,1-Trichloroethane	9,080	O	94	O	9,174
Tinker	Chlorodifluoromethane	7,469	O	0	O	7,469
Tinker	1,2-Dichlorobenzene	343	O	6,326	O	6,669
Tinker	Chromium Compounds	654	O	4,611	O	5,265
Tinker	Methyl Isobutyl Ketone	341	O	4,020	O	4,361
Tinker	Nickel	1,297	O	77	O	1,374
Tinker	Cobalt	728	O	83	O	811
Tinker	Dichlorodifluoromethane	245	O			245
Tinker	Manganese	15	O	1	O	16
Tinker	Copper	5	O	0	O	5
Tinker	Nitric Acid	0	O	0	O	0
Tinker	Phosphoric Acid	0	O	0	O	0
Tinker	Sulfuric Acid	0	O	0	O	0
						0
UTTR	Nitroglycerin					0
						0
WPAFB	Hydrochloric Acid	0	O	120,000	O	120,000
WPAFB	Sulfuric Acid	0	O	28,000	O	28,000
						0
McChord	Methyl Ethyl Ketone	17,000	C			17,000
						0
Travis	Methyl Ethyl Ketone	6,700	C			6,700
						0
Columbus	Methyl Ethyl Ketone	3,000	C			3,000
Columbus	Chlorine	42	E			42
						0
Randolph	Methyl Ethyl Ketone	2,427	C	13,327	C	15,754
						0
Sheppard	Chlorine	245	O			245
						0
Tyndall	Toluene	62	E			62
Tyndall	Chlorine	57	E	0	O	57
Tyndall	Xylene (mixed isomers)	6	E			6
Tyndall	Methyl Tert-Butyl Ether	0	O			0
						0
Bolling	Hydroquinone					0
						0
Academy	Toluene	4,500	C			4,500
Academy	Xylene (mixed isomers)	750	C			750
						0
Beale	Chlorine					0
						0
Cannon	Methyl Ethyl Ketone	13,927	O			13,927
Cannon	Toluene	12,700	O			12,700
						0
Castle	Chlorine					0
						0
Davis Monthan	Methyl Ethyl Ketone	16,146	C			16,146
Davis Monthan	Chlorine					0
						0
Ellsworth	Chlorine	11,000	O			11,000
						0
Little Rock	Methyl Ethyl Ketone	32	O	22,559	O	22,591
						0
Nellis	Methyl Ethyl Ketone	10,559	E			10,559
						0
Shaw	Methyl Ethyl Ketone	11,687	O			11,687
						0
Elmendorf	Methyl Ethyl Ketone	11,501	E			11,501
Elmendorf	Sulfuric Acid					0
Elmendorf	Ammonia					0
Elmendorf	Bromochlorodifluoromethane					0

<b>CY 94 TRI Releases to Surface Water</b>			
<b>Installation</b>	<b>Chemical/Chemical Category</b>	<b>Lbs Released</b>	<b>BOE</b>
Arnold	Sulfuric Acid	326	M
Hill	Ethylene Glycol	750	O
Kelly	Phenol	85	M
McClellan	Dichloromethane	0	M
McClellan	Methyl Ethyl Ketone	0	M
McClellan	Toluene	0	M
McClellan	Methyl Isobutyl Ketone	0	M
McClellan	Tetrachloroethylene	0	M
McClellan	Zinc Compounds	0	M
Robins	Dichloromethane	0	M
Robins	1,1,1-Trichloroethane	102	M
Robins	Phenol	43	M
Tinker	Dichloromethane	0	O
Tinker	Phenol	1	M
Tinker	Methyl Ethyl Ketone	11	O
Tinker	Glycol Ethers	0	O
Tinker	Tetrachloroethylene	0	O
Tinker	Toluene	0	O
Tinker	Xylene (mixed isomers)	0	O
Tinker	1,1,1-Trichloroethane	0	O
Tinker	1,2-Dichlorobenzene	0	O
Tinker	Chromium Compounds	173	M
Tinker	Nickel	121	O
Tinker	Cobalt	0	O
Tinker	Manganese	63	O
Tinker	Copper	48	O
Tinker	Nitric Acid	0	M
Tinker	Phosphoric Acid	0	M
Tinker	Sulfuric Acid	0	M
Columbus	Methyl Ethyl Ketone	0	O
Columbus	Chlorine	0	O
Sheppard	Chlorine	0	M
Beale	Chlorine	9,000	C
Castle	Chlorine	0	C
Ellsworth	Chlorine	3,200	M



<b><u>CY 94 TRI Releases to Land</u></b>				
<b>Installation</b>	<b>Chemical/Chemical Category</b>	<b>Releases to Land</b>	<b>BOE</b>	<b>Type release</b>
Beale	Chlorine	2932	C	Landfill
Edwards	Glycol Ethers	12430	M	Surface Impoundment
Elmendorf	Ammonia	45000	M	Other
Hill	Ethylene Glycol	4700	O	Other
Shaw	Methyl Ethyl Ketone	615	O	Other

## Appendix B: Telephone Questionnaire

### METHOD

Did you perform your TRI process in-house or by contract?

In-house      Contract

### TRAINING

Did you receive the AF-sponsored EPCRA training put on by GAIA Corp.?

Yes      No

### TRAINING HELPFULNESS

If yes, how helpful was the training to your TRI reporting process?

Very      Somewhat      Not at all

### START DATE

When did you start the TRI process?

### PHARMACY

Did your base have a Hazardous Material Pharmacy up and running during CY 94?

Yes      No

### PRIMARY DATABASE

What database did you primarily use to determine if the threshold for a particular chemical was exceeded?

CEMAS      SBSS      Pharmacy      BEE      Other

### COMPLETENESS OF DATA

Was the data complete or did it have obvious holes?

Complete      Holes

If data was lacking, how did you estimate material use?

### PRIMARY GUIDANCE

What guidance document did you primarily use to help your TRI process?

AF   DoD   EO 12856   EPCRA   GAIA Manual   GAIA Handbook   40 CFR   State

### LIST OF PROCESSES

Did you develop a list of processes that generate TRI chemicals?

Yes      No

### PROCESS VISITATION

Did you visit the process sites?

Yes      No

## Appendix C: Raw Survey Data

Installation	Method	Training	Helpful	Start	Pharmacy?	Database	Complete?	Guidance	List	Visit	Command
Academy	C	N		Mar-95	N	BEE	N	FORM R	N	Y	NONE
Arnold	C	Y	VERY	Apr-95	YES	PHARM	Y	GAIA MAN	N	Y	NONE
Beale	I	Y	VERY	UNK	YES	SBSS	Y	EPA	Y	Y	ACC
Bolling	C	Y	SW	1ST QTR	NO	BEE	N	AF	Y	Y	NONE
Cannon	I	N		May-95	YES	BEE	N	FORM R	Y	Y	ACC
Castle	I	N		Jan-95	N	BEE	Y	FORM R	N	N	ACC
Charleston	C	Y	VERY	Dec-94	YES	BEE	N	GAIA MAN	Y	Y	AMC
Columbus	C	Y	SW	May-95	NO	SBSS	N	AF	Y	Y	AETC
Davis Monthan	I	Y	SW	Feb-95	NO	CEMAS	N	GAIA HAND	N	N	ACC
Dover	C	N		Jan-95	YES	SBSS	N	FORM R	N	Y	AMC
Edwards	C	Y	VERY	Apr-95	NO	CONTRAC	N	GAIA MAN	N	Y	AFMC
Eglin	C	Y	VERY	Jun-94	NO	SBSS	N	STATE	N	Y	AFMC
Ellsworth	I	Y	VERY	Apr-95	YES	BEE	N	AF	Y	N	ACC
Elmendorf	I	Y	VERY	1ST QTR	YES	BEE	Y	DoD	N	N	PACAF
Goodfellow	C	N		Apr-95	NO	SBSS	N	AF	N	Y	AETC
Grand Forks	C	Y	SW	Jan-95	NO	SBSS	N	AF	Y	Y	AMC
Hanscom	I	Y	VERY	Mar-95	YES	BEE	N	GAIA HAND	N	N	AFMC
Hill	C	Y	SW	Nov-94	YES	BEE	N	AF	Y	Y	AFMC
Kelly	I	N		FALL 93	YES	PHARM	Y	AF	Y	Y	AFMC
Little Rock	I	Y	VERY	Mar-95	NO	BEE	Y	GAIA MAN	N	Y	ACC
McChord	C	N		1ST QTR	YES	SURVEY	N	AF	Y	Y	AMC
McClellan	C	N		FALL 93	NO	SBSS	Y	AF	Y	Y	AFMC
Moody	I	Y	SW	1ST QTR	NO	SBSS	N	AF	N	Y	ACC
Nellis	I	Y	VERY	FALL 93	YES	BEE	N	GAIA MAN	N	Y	ACC
Randolph	C	Y	SW	Feb-95	NO	SBSS	N	AF	Y	Y	AETC
Reese	C	N		Feb-95	NO	SBSS	N	AQF	Y	Y	AETC
Robins	C	Y	VERY	Oct-93	NO	BEE	N	FORM R	Y	Y	AFMC
S-J	I	N		May-95	NO	BEE	N	AF	Y	N	ACC
Shaw	I	Y	SW	Mar-95	NO	BEE	N	FORM R	N	N	ACC
Sheppard	C	N		Apr-95	NO	SBSS	N	EPA	Y	Y	AETC
Tinker	I	Y	VERY	MID 94	YES	PHARM	Y	FORM R	Y	Y	AFMC
Travis	C	N		Feb-95	NO	SBSS	N	AF	Y	Y	AMC
Tyndall	C	N		Apr-95	N	SBSS	N	STATE	Y	Y	AETC
UTTR	C	Y	SW	Nov-94	YES	BEE	N	AF	Y	Y	AFMC
Whiteman	I	Y	VERY	Jan-95	NO	BEE	N	GAIA HAND	Y	Y	ACC
WPAFB	C	Y	VERY	Jan-95	NO	BEE	N	AF	N	Y	AFMC

### Appendix D: Tally Sheets

<b><u>Reported?</u></b>	<b><u>Method</u></b>	
	Inhouse	Contract
Yes	11	15
No	4	6

<b><u>Reported?</u></b>	<b><u>Training</u></b>	
	Yes	No
Yes	9	17
No	4	6

<b><u>Reported?</u></b>	<b><u>Training Helpfulness</u></b>		
	Very	Somewhat	Not at all
Yes	10	7	0
No	4	2	0

<b><u>Reported?</u></b>	<b><u>Start Date</u></b>			
	Pre-95	1st Qtr 95	2nd Qtr 95	Unknown
Yes	6	10	8	2
No	2	6	2	0

<b><u>Reported?</u></b>	<b><u>Pharmacy?</u></b>	
	Yes	No
Yes	12	14
No	3	7

<b><u>Reported?</u></b>	<b><u>Data Complete?</u></b>	
	Yes	No
Yes	8	18
No	0	10

<b><u>Reported?</u></b>	<b><u>Database</u></b>			
	SBSS	Pharmacy	Bio	Other
Yes	7	3	13	3
No	6	0	4	0

<b><u>Reported?</u></b>	<b><u>Guidance</u></b>						
	AF	DoD	GAIA Manual	GAIA Handbook	EPA	State	Form R
Yes	11	1	4	1	2	1	6
No	5	0	1	2	0	1	1

<u>Reported?</u>	<u>List of Processes?</u>	
	Yes	No
Yes	16	10
No	5	5

<u>Reported?</u>	<u>Visited Sites?</u>	
	Yes	No
Yes	21	5
No	8	2

## Appendix E: Statistical Analysis (Mathcad Templates)



**(1) State Focus Question to be Answered**

Is the facility's status of reporting (Report/Did Not Report) related to the method used to develop the TRI data (Inhouse vs. Contract)?

**(2) State Assumptions**

Assumptions:

1. Each observation is classified into exactly one cell.
2. The observations are observations on a random sample. Each observation has the same probability of being classified into cell (r,c) as any other observation.
3. The row totals are given, not random.

**(3) State Hypothesis**

$H_0$ : The populations of Report/Did Not Report are homogeneous with respect to the method used for the TRI process (Inhouse vs. Contract)

$H_a$ : The populations of Report/Did Not Report are not homogeneous with respect to the method used for the TRI process (Inhouse vs. Contract)

**(4) Specify a Test Statistic**

$$\chi^2 = \sum_r \sum_c \frac{(O_{r,c} - E_{r,c})^2}{E_{r,c}}$$

where  $r$  and  $c$  run over the rows and columns of the observed and expected value matrices.

The P-value will be computed as

$$Pvalue = 1 - \text{chisq}(df, \chi^2)$$

with  $df$  defined as  $df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1)$

**(5) State the Distribution of the Test Statistic**

Under the null hypothesis, for a reasonable sample size, the statistic  $\chi^2$  computed above will approximately follow a  $\chi^2$  distribution having degrees of freedom

$$df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1)$$

**(6) Formulate a Decision Rule**

If  $Pvalue > \alpha$  then do not reject  $H_0$  in favor of  $H_a$  at significance level  $\alpha$

**(7) Obtain a Random Sample**

$$\text{method} = \begin{matrix} & \begin{matrix} \text{In} & \text{Cont} \end{matrix} \\ \begin{pmatrix} 11 & 15 \\ 4 & 6 \end{pmatrix} & \begin{matrix} \text{Reported} \\ \text{Did Not Report} \end{matrix} \end{matrix}$$

**(8) Compute the Observed Value of the Test Statistic**

$$O = \text{method}$$

$$r = 0 \dots \text{rows}(O) - 1 \quad c = 0 \dots \text{cols}(O) - 1$$

$$N = \sum_r \sum_c \text{method}_{r,c} \quad N = 36$$

$$PC_c = \frac{\sum O^{<c>}}{N} \quad PC^T = (0.417 \quad 0.583) \quad \sum PC = 1$$

$$PR_r = \frac{\sum (O^T)^{<r>}}{N} \quad PR = \begin{pmatrix} 0.722 \\ 0.278 \end{pmatrix} \quad \sum PR = 1$$

Under the assumption of independence between column and row categories, the probability of being classified into the (r,c) cell will be

$$PR_r \cdot PC_c$$

the product of the corresponding row and column probabilities. A matrix of these probabilities can then be built.

$$P_{r,c} = PR_r \cdot PC_c \quad P = \begin{pmatrix} 0.301 & 0.421 \\ 0.116 & 0.162 \end{pmatrix} \quad \sum_r \sum_c P_{r,c} = 1$$

The expected cell values are then given by  $E = N \cdot P$

$$E = \begin{pmatrix} 10.833 & 15.167 \\ 4.167 & 5.833 \end{pmatrix}$$

and the test statistic can then easily be computed

$$\chi^2 = \sum_r \sum_c \frac{(O_{r,c} - E_{r,c})^2}{E_{r,c}} \quad \chi^2 = 0.016$$

The degrees of freedom associated with this test statistic are:

$$df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1) \quad df = 1$$

The P-value is then calculated...

$$Pvalue = 1 - \text{chisq}(df, \chi^2) \quad Pvalue = 0.9 \quad \alpha = .1$$

**(9) Use the Decision Rule to state a Knowledge Claim**

$Pvalue > \alpha = 1$  indicates that  $Pvalue > \alpha$ , therefore do not reject the null hypothesis.

**(10) Consider the Practical Significance and State a Value Claim**

At a significance level of .1, the statistical analysis of the contingency table indicates that it is not possible to reject the claim that homogeneity exists across the data set.

**(11) Use the Knowledge Claim and the Value Claim to Answer the Focus Question**

The facility's status of reporting (Report/Did Not Report) does not appear to be related to the method used to develop the TRI data (Inhouse vs. Contract).

**(1) State Focus Question to be Answered**

Is the facility's status of reporting (Report/Did Not Report) related to whether the TRI POC received the GAIA training?

**(2) State Assumptions**

Assumptions:

1. Each observation is classified into exactly one cell.
2. The observations are observations on a random sample. Each observation has the same probability of being classified into cell (r,c) as any other observation.
3. The row totals are given, not random.

**(3) State Hypothesis**

**H<sub>0</sub>:** The populations of Report/Did Not Report are homogeneous with respect to whether the TRI POC received the GAIA training?

**H<sub>a</sub>:** The populations of Report/Did Not Report are not homogeneous with respect to whether the TRI POC received the GAIA training?

**(4) Specify a Test Statistic**

$$\chi^2 = \sum_r \sum_c \frac{(O_{r,c} - E_{r,c})^2}{E_{r,c}}$$

where **r** and **c** run over the rows and columns of the observed and expected value matrices.

The P-value will be computed as

$$\text{Pvalue} = 1 - \text{chisq}(\text{df}, \chi^2)$$

with df defined as  $\text{df} = (\text{rows}(\text{O}) - 1) \cdot (\text{cols}(\text{O}) - 1)$

**(5) State the Distribution of the Test Statistic**

Under the null hypothesis, for a reasonable sample size, the statistic  $\chi^2$  computed above will approximately follow a  $\chi^2$  distribution having degrees of freedom

$$\text{df} = (\text{rows}(\text{O}) - 1) \cdot (\text{cols}(\text{O}) - 1)$$

**(6) Formulate a Decision Rule**

If  $\text{Pvalue} > \alpha$  then do not reject  $H_0$  in favor of  $H_a$  at significance level  $\alpha$

**(7) Obtain a Random Sample**

$$\text{training} = \begin{matrix} & \begin{matrix} \text{Yes} & \text{No} \end{matrix} \\ \begin{pmatrix} 9 & 17 \\ 4 & 6 \end{pmatrix} & \begin{matrix} \text{Reported} \\ \text{Did Not Report} \end{matrix} \end{matrix}$$

**(8) Compute the Observed Value of the Test Statistic**

$$O = \text{training}$$

$$r = 0 \dots \text{rows}(O) - 1 \quad c = 0 \dots \text{cols}(O) - 1$$

$$N = \sum_r \sum_c \text{training}_{r,c} \quad N = 36$$

$$PC_c = \frac{\sum O^{<c>}}{N} \quad PC^T = (0.361 \quad 0.639) \quad \sum PC = 1$$

$$PR_r = \frac{\sum (O^T)^{<r>}}{N} \quad PR = \begin{pmatrix} 0.722 \\ 0.278 \end{pmatrix} \quad \sum PR = 1$$

Under the assumption of independence between column and row categories, the probability of being classified into the (r,c) cell will be

$$PR_r \cdot PC_c$$

the product of the corresponding row and column probabilities. A matrix of these probabilities can then be built as follows:

$$P_{r,c} = PR_r \cdot PC_c \quad P = \begin{pmatrix} 0.261 & 0.461 \\ 0.1 & 0.177 \end{pmatrix} \quad \sum_r \sum_c P_{r,c} = 1$$

The expected cell values are then given by  $E = N \cdot P$

$$E = \begin{pmatrix} 9.389 & 16.611 \\ 3.611 & 6.389 \end{pmatrix}$$

and the test statistic can then easily be computed

$$\chi^2 = \sum_r \sum_c \frac{(O_{r,c} - E_{r,c})^2}{E_{r,c}} \quad \chi^2 = 0.091$$

The degrees of freedom associated with this test statistic are:

$$df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1) \quad df = 1$$

The P-value is then calculated...

$$Pvalue = 1 - \text{chisq}(df, \chi^2) \quad Pvalue = 0.763 \quad \alpha = .1$$

**(9) Use the Decision Rule to state a Knowledge Claim**

$Pvalue > \alpha = 1$  indicates that  $Pvalue > \alpha$ , therefore do not reject the null hypothesis.

**(10) Consider the Practical Significance and State a Value Claim**

At a significance level of .1, the statistical analysis of the contingency table indicates that it is not possible to reject the claim that homogeneity exists across the data set.

**(11) Use the Knowledge Claim and the Value Claim to Answer the Focus Question**

The facility's status of reporting (Report/Did Not Report) does not appear to be related to whether the TRI POC received the GAIA training.

**(1) State Focus Question to be Answered**

For the 23 facilities whose TRI POCs received the GAIA training, is the facility's status of reporting (Report/Did Not Report) related to whether they found the training to be very or just somewhat helpful?

**(2) State Assumptions**

Assumptions:

1. Each observation is classified into exactly one cell.
2. The observations are observations on a random sample. Each observation has the same probability of being classified into cell (r,c) as any other observation.
3. The row totals are given, not random.

**(3) State Hypothesis**

$H_0$ : The populations of Report/Did Not Report are homogeneous with respect to whether the TRI POCs found the GAIA training very or just somewhat helpful

$H_a$ : The populations of Report/Did Not Report are not homogeneous with respect to whether the TRI POCs found the GAIA training very or just somewhat helpful

**(4) Specify a Test Statistic**

$$\chi^2 = \sum_r \sum_c \frac{(O_{r,c} - E_{r,c})^2}{E_{r,c}}$$

where  $r$  and  $c$  run over the rows and columns of the observed and expected value matrices.

The P-value will be computed as

$$Pvalue = 1 - \text{chisq}(df, \chi^2)$$

with df defined as  $df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1)$

**(5) State the Distribution of the Test Statistic**

Under the null hypothesis, for a reasonable sample size, the statistic  $\chi^2$  computed above will approximately follow a  $\chi^2$  distribution having degrees of freedom

$$df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1)$$

**(6) Formulate a Decision Rule**

If  $Pvalue > \alpha$  then do not reject  $H_0$  in favor of  $H_a$  at significance level  $\alpha$

**(7) Obtain a Random Sample**

Very SW

$$\text{helpful} := \begin{pmatrix} 10 & 7 \\ 4 & 2 \end{pmatrix} \begin{matrix} \text{Reported} \\ \text{Did Not Report} \end{matrix}$$

**(8) Compute the Observed Value of the Test Statistic**

$$O := \text{helpful}$$

$$r := 0 \dots \text{rows}(O) - 1 \quad c := 0 \dots \text{cols}(O) - 1$$

$$N := \sum_r \sum_c \text{helpful}_{r,c} \quad N = 23$$

$$PC_c = \frac{\sum O^{<c>}}{N} \quad PC^T = (0.609 \quad 0.391) \quad \sum PC = 1$$

$$PR_r = \frac{\sum (O^T)^{<r>}}{N} \quad PR = \begin{pmatrix} 0.739 \\ 0.261 \end{pmatrix} \quad \sum PR = 1$$

Under the assumption of independence between column and row categories, the probability of being classified into the (r,c) cell will be

$$PR_r \cdot PC_c$$

the product of the corresponding row and column probabilities. A matrix of these probabilities can be built as follows:

$$P_{r,c} := PR_r \cdot PC_c \quad P = \begin{pmatrix} 0.45 & 0.289 \\ 0.159 & 0.102 \end{pmatrix} \quad \sum_r \sum_c P_{r,c} = 1$$

The expected cell values are then given by  $E := N \cdot P$

$$E = \begin{pmatrix} 10.348 & 6.652 \\ 3.652 & 2.348 \end{pmatrix}$$



and the test statistic can then easily be computed

$$\chi^2 = \sum_r \sum_c \frac{(O_{r,c} - E_{r,c})^2}{E_{r,c}} \quad \chi^2 = 0.115$$

The degrees of freedom associated with this test statistic are:

$$df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1) \quad df = 1$$

The P-value is then calculated...

$$Pvalue = 1 - \text{chisq}(df, \chi^2) \quad Pvalue = 0.735 \quad \alpha = .1$$

**(9) Use the Decision Rule to state a Knowledge Claim**

$Pvalue > \alpha = 1$  indicates that  $Pvalue > \alpha$ , therefore do not reject the null hypothesis.

**(10) Consider the Practical Significance and State a Value Claim**

At a significance level of .1, the statistical analysis of the contingency table indicates that it is not possible to reject the claim that homogeneity exists across the data set.

**(11) Use the Knowledge Claim and the Value Claim to Answer the Focus Question**

For the 23 facilities whose TRI POCs received the GAIA training, the facility's status of reporting (Report/Did Not Report) does not appear to be related to whether they found the training to be very or just somewhat helpful.

**(1) State Focus Question to be Answered**

Is the facility's status of reporting (Report/Did Not Report) related to when they started their TRI reporting process?

**(2) State Assumptions**

Assumptions:

1. Each observation is classified into exactly one cell.
2. The observations are observations on a random sample. Each observation has the same probability of being classified into cell (r,c) as any other observation.
3. The row totals are given, not random.

**(3) State Hypothesis**

$H_0$ : The populations of Report/Did Not Report are homogeneous with respect to when they started their TRI reporting process?

$H_a$ : The populations of Report/Did Not Report are not homogeneous with respect to when they started their TRI reporting process?

**(4) Specify a Test Statistic**

$$\chi^2 = \sum_r \sum_c \frac{(O_{r,c} - E_{r,c})^2}{E_{r,c}}$$

where r and c run over the rows and columns of the observed and expected value matrices.

The P-value will be computed as

$$Pvalue = 1 - \text{chisq}(df, \chi^2)$$

with df defined as  $df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1)$

**(5) State the Distribution of the Test Statistic**

Under the null hypothesis, for a reasonable sample size, the statistic  $\chi^2$  computed above will approximately follow a  $\chi^2$  distribution having degrees of freedom

$$df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1)$$

**(6) Formulate a Decision Rule**

If  $Pvalue > \alpha$  then do not reject  $H_0$  in favor of  $H_a$  at significance level  $\alpha$

**(7) Obtain a Random Sample**

$$\text{start} = \begin{pmatrix} 6 & 10 & 8 & 2 \\ 2 & 6 & 2 & 0 \end{pmatrix} \begin{array}{l} \text{Reported} \\ \text{Did Not Report} \end{array}$$

The categories from left to right are: Pre-95, Jan-Mar 95, Apr - Jun 95, and Unknown

**(8) Compute the Observed Value of the Test Statistic**

$$O = \text{start}$$

$$r = 0 \dots \text{rows}(O) - 1 \quad c = 0 \dots \text{cols}(O) - 1$$

$$N = \sum_r \sum_c \text{start}_{r,c} \quad N = 36$$

$$PC_c = \frac{\sum O^{<c>}}{N} \quad PC^T = (0.222 \quad 0.444 \quad 0.278 \quad 0.056) \quad \sum PC = 1$$

$$PR_r = \frac{\sum (O^T)^{<r>}}{N} \quad PR = \begin{pmatrix} 0.722 \\ 0.278 \end{pmatrix} \quad \sum PR = 1$$

Under the assumption of independence between column and row categories, the probability of being classified into the (r,c) cell will be

$$PR_r \cdot PC_c$$

the product of the corresponding row and column probabilities. A matrix of these probabilities can be built as follows:

$$P_{r,c} = PR_r \cdot PC_c \quad P = \begin{pmatrix} 0.16 & 0.321 & 0.201 & 0.04 \\ 0.062 & 0.123 & 0.077 & 0.015 \end{pmatrix} \quad \sum_r \sum_c P_{r,c} = 1$$

The expected cell values are then given by  $E = N \cdot P$

$$E = \begin{pmatrix} 5.778 & 11.556 & 7.222 & 1.444 \\ 2.222 & 4.444 & 2.778 & 0.556 \end{pmatrix}$$

and the test statistic can then easily be computed

$$\chi^2 = \sum_r \sum_c \frac{(O_{r,c} - E_{r,c})^2}{E_{r,c}} \quad \chi^2 = 1.855$$

The degrees of freedom associated with this test statistic are:

$$df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1) \quad df = 3$$

The P-value is then calculated...

$$Pvalue = 1 - \text{chisq}(df, \chi^2) \quad Pvalue = 0.603 \quad \alpha = .1$$

**(9) Use the Decision Rule to state a Knowledge Claim**

$Pvalue > \alpha = 1$  indicates that  $Pvalue > \alpha$ , therefore do not reject the null hypothesis.

**(10) Consider the Practical Significance and State a Value Claim**

At a significance level of .1, the statistical analysis of the contingency table indicates that it is not possible to reject the claim that homogeneity exists across the data set.

**(11) Use the Knowledge Claim and the Value Claim to Answer the Focus Question**

The facility's status of reporting (Report/Did Not Report) does not appear to be related to when they started their TRI reporting process.

**(1) State Focus Question to be Answered**

Is the facility's status of reporting (Report/Did Not Report) related to whether their facility had a Hazardous Material Pharmacy up and running during calendar year 94?

**(2) State Assumptions**

Assumptions:

1. Each observation is classified into exactly one cell.
2. The observations are observations on a random sample. Each observation has the same probability of being classified into cell  $(r,c)$  as any other observation.
3. The row totals are given, not random.

**(3) State Hypothesis**

$H_0$ : The populations of Report/Did Not Report are homogeneous with respect to whether the facility had a Hazardous Material Pharmacy up and running during CY 94

$H_a$ : The populations of Report/Did Not Report are not homogeneous with respect to whether the facility had a Hazardous Material Pharmacy up and running during CY 94

**(4) Specify a Test Statistic**

$$\chi^2 = \sum_r \sum_c \frac{(O_{r,c} - E_{r,c})^2}{E_{r,c}}$$

where  $r$  and  $c$  run over the rows and columns of the observed and expected value matrices.

The P-value will be computed as

$$Pvalue = 1 - \text{chisq}(df, \chi^2)$$

with  $df$  defined as  $df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1)$

**(5) State the Distribution of the Test Statistic**

Under the null hypothesis, for a reasonable sample size, the statistic  $\chi^2$  computed above will approximately follow a  $\chi^2$  distribution having degrees of freedom

$$df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1)$$

**(6) Formulate a Decision Rule**

If  $Pvalue > \alpha$  then do not reject  $H_0$  in favor of  $H_a$  at significance level  $\alpha$

**(7) Obtain a Random Sample**

$$\text{pharmacy} := \begin{matrix} & \begin{matrix} \text{Yes} & \text{No} \end{matrix} \\ \begin{pmatrix} 12 & 14 \\ 3 & 7 \end{pmatrix} & \begin{matrix} \text{Reported} \\ \text{Did Not Report} \end{matrix} \end{matrix}$$

**(8) Compute the Observed Value of the Test Statistic**

$$O := \text{pharmacy}$$

$$r := 0 \dots \text{rows}(O) - 1 \quad c := 0 \dots \text{cols}(O) - 1$$

$$N := \sum_r \sum_c \text{pharmacy}_{r,c} \quad N = 36$$

$$PC_c := \frac{\sum O^{<c>}}{N} \quad PC^T = (0.417 \quad 0.583) \quad \sum PC = 1$$

$$PR_r := \frac{\sum (O^T)^{<r>}}{N} \quad PR = \begin{pmatrix} 0.722 \\ 0.278 \end{pmatrix} \quad \sum PR = 1$$

Under the assumption of independence between column and row categories, the probability of being classified into the (r,c) cell will be

$$PR_r \cdot PC_c$$

the product of the corresponding row and column probabilities. A matrix of these probabilities can then be built.

$$P_{r,c} := PR_r \cdot PC_c \quad P = \begin{pmatrix} 0.301 & 0.421 \\ 0.116 & 0.162 \end{pmatrix} \quad \sum_r \sum_c P_{r,c} = 1$$

The expected cell values are then given by  $E := N \cdot P$

$$E = \begin{pmatrix} 10.833 & 15.167 \\ 4.167 & 5.833 \end{pmatrix}$$

and the test statistic can then easily be computed

$$\chi^2 = \sum_r \sum_c \frac{(O_{r,c} - E_{r,c})^2}{E_{r,c}} \quad \chi^2 = 0.775$$

The degrees of freedom associated with this test statistic are:

$$df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1) \quad df = 1$$

The P-value is then calculated...

$$P\text{value} = 1 - \text{chisq}(df, \chi^2) \quad P\text{value} = 0.379 \quad \alpha = .1$$

**(9) Use the Decision Rule to state a Knowledge Claim**

$P\text{value} > \alpha = 1$  indicates that  $P\text{value} > \alpha$ , therefore do not reject the null hypothesis.

**(10) Consider the Practical Significance and State a Value Claim**

At a significance level of .1, the statistical analysis of the contingency table indicates that it is not possible to reject the claim that homogeneity exists across the data set.

**(11) Use the Knowledge Claim and the Value Claim to Answer the Focus Question**

The facility's status of reporting (Report/Did Not Report) does not appear to be related to whether or not their facility had a Hazardous Material Pharmacy up and running during calendar year 94.

**(1) State Focus Question to be Answered**

Is the facility's status of reporting (Report/Did Not Report) related to which database was primarily used to determine if the threshold for any particular chemical was exceeded?

**(2) State Assumptions**

Assumptions:

1. Each observation is classified into exactly one cell.
2. The observations are observations on a random sample. Each observation has the same probability of being classified into cell (r,c) as any other observation.
3. The row totals are given, not random.

**(3) State Hypothesis**

**H<sub>0</sub>:** The populations of Report/Did Not Report are homogeneous with respect to which database was primarily used to determine if the threshold for any particular chemical was exceeded

**H<sub>a</sub>:** The populations of Report/Did Not Report are not homogeneous with respect to which database was primarily used to determine if the threshold for any particular chemical was exceeded

**(4) Specify a Test Statistic**

$$\chi^2 = \sum_r \sum_c \frac{(O_{r,c} - E_{r,c})^2}{E_{r,c}}$$

where **r** and **c** run over the rows and columns of the observed and expected value matrices.

The P-value will be computed as

$$Pvalue = 1 - \text{chisq}(df, \chi^2)$$

with df defined as  $df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1)$

**(5) State the Distribution of the Test Statistic**

Under the null hypothesis, for a reasonable sample size, the statistic  $\chi^2$  computed above will approximately follow a  $\chi^2$  distribution having degrees of freedom

$$df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1)$$

**(6) Formulate a Decision Rule**

If  $Pvalue > \alpha$  then do not reject  $H_0$  in favor of  $H_a$  at significance level  $\alpha$



### (7) Obtain a Random Sample

$$\text{database} = \begin{pmatrix} 7 & 3 & 13 & 3 \\ 6 & 0 & 4 & 0 \end{pmatrix} \begin{matrix} \text{Reported} \\ \text{Did Not Report} \end{matrix}$$

The categories from left to right are SBSS, Pharmacy, Bio, and Other (includes CEMAS and other databases)

### (8) Compute the Observed Value of the Test Statistic

$O = \text{database}$

$r = 0 \dots \text{rows}(O) - 1$   $c = 0 \dots \text{cols}(O) - 1$

$$N = \sum_r \sum_c \text{database}_{r,c} \quad N = 36$$

$$PC_c = \frac{\sum O^{<c>}}{N} \quad PC^T = (0.361 \quad 0.083 \quad 0.472 \quad 0.083) \quad \sum PC = 1$$

$$PR_r = \frac{\sum (O^T)^{<r>}}{N} \quad PR = \begin{pmatrix} 0.722 \\ 0.278 \end{pmatrix} \quad \sum PR = 1$$

Under the assumption of independence between column and row categories, the probability of being classified into the  $(r,c)$  cell will be

$$PR_r \cdot PC_c$$

the product of the corresponding row and column probabilities. A matrix of these probabilities can then be built as follows:

$$P_{r,c} = PR_r \cdot PC_c \quad P = \begin{pmatrix} 0.261 & 0.06 & 0.341 & 0.06 \\ 0.1 & 0.023 & 0.131 & 0.023 \end{pmatrix} \quad \sum_r \sum_c P_{r,c} = 1$$

The expected cell values are then given by  $E = N \cdot P$

$$E = \begin{pmatrix} 9.389 & 2.167 & 12.278 & 2.167 \\ 3.611 & 0.833 & 4.722 & 0.833 \end{pmatrix}$$

and the test statistic can then easily be computed

$$\chi^2 = \sum_r \sum_c \frac{(O_{r,c} - E_{r,c})^2}{E_{r,c}} \quad \chi^2 = 4.649$$

The degrees of freedom associated with this test statistic are:

$$df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1) \quad df = 3$$

The P-value is then calculated...

$$Pvalue = 1 - \text{chisq}(df, \chi^2) \quad Pvalue = 0.199 \quad \alpha = .1$$

**(9) Use the Decision Rule to state a Knowledge Claim**

$Pvalue > \alpha = 1$  indicates that  $Pvalue > \alpha$ , therefore do not reject the null hypothesis.

**(10) Consider the Practical Significance and State a Value Claim**

At a significance level of .1, the statistical analysis of the contingency table indicates that it is not possible to reject the claim that homogeneity exists across the data set.

**(11) Use the Knowledge Claim and the Value Claim to Answer the Focus Question**

The facility's status of reporting (Report/Did Not Report) does not appear to be related to which database was primarily used to determine if the threshold for any particular chemical was exceeded

**(1) State Focus Question to be Answered**

Is the facility's status of reporting (Report/Did Not Report) related to whether the database used was the Standard Base Supply System versus some other system?

**(2) State Assumptions**

Assumptions:

1. Each observation is classified into exactly one cell.
2. The observations are observations on a random sample. Each observation has the same probability of being classified into cell (r,c) as any other observation.
3. The row totals are given, not random.

**(3) State Hypothesis**

$H_0$ : The populations of Report/Did Not Report are homogeneous with respect to the database primarily used (SBSS vs. Non-SBSS)

$H_a$ : The populations of Report/Did Not Report are not homogeneous with respect to the database primarily used (SBSS vs. Non-SBSS)

**(4) Specify a Test Statistic**

$$\chi^2 = \sum_r \sum_c \frac{(O_{r,c} - E_{r,c})^2}{E_{r,c}}$$

where  $r$  and  $c$  run over the rows and columns of the observed and expected value matrices.

The P-value will be computed as

$$Pvalue = 1 - \text{chisq}(df, \chi^2)$$

with  $df$  defined as  $df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1)$

**(5) State the Distribution of the Test Statistic**

Under the null hypothesis, for a reasonable sample size, the statistic  $\chi^2$  computed above will approximately follow a  $\chi^2$  distribution having degrees of freedom

$$df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1)$$

**(6) Formulate a Decision Rule**

If  $Pvalue > \alpha$  then do not reject  $H_0$  in favor of  $H_a$  at significance level  $\alpha$

**(7) Obtain a Random Sample**

$$\begin{array}{cc} \text{SBSS} & \text{Non-SBSS} \\ \text{SBSS} = \begin{pmatrix} 7 & 19 \\ 6 & 4 \end{pmatrix} & \begin{array}{l} \text{Reported} \\ \text{Did Not Report} \end{array} \end{array}$$

**(8) Compute the Observed Value of the Test Statistic**

$$O = \text{SBSS}$$

$$r = 0 \dots \text{rows}(O) - 1 \quad c = 0 \dots \text{cols}(O) - 1$$

$$N = \sum_r \sum_c \text{SBSS}_{r,c} \quad N = 36$$

$$\text{PC}_c = \frac{\sum O^{<c>}}{N} \quad \text{PC}^T = (0.361 \quad 0.639) \quad \sum \text{PC} = 1$$

$$\text{PR}_r = \frac{\sum (O^T)^{<r>}}{N} \quad \text{PR} = \begin{pmatrix} 0.722 \\ 0.278 \end{pmatrix} \quad \sum \text{PR} = 1$$

Under the assumption of independence between column and row categories, the probability of being classified into the (r,c) cell will be

$$\text{PR}_r \cdot \text{PC}_c$$

the product of the corresponding row and column probabilities. A matrix of these probabilities can then be built as follows:

$$P_{r,c} = \text{PR}_r \cdot \text{PC}_c \quad P = \begin{pmatrix} 0.261 & 0.461 \\ 0.1 & 0.177 \end{pmatrix} \quad \sum_r \sum_c P_{r,c} = 1$$

The expected cell values are then given by  $E = N \cdot P$

$$E = \begin{pmatrix} 9.389 & 16.611 \\ 3.611 & 6.389 \end{pmatrix}$$

and the test statistic can then easily be computed

$$\chi^2 = \sum_r \sum_c \frac{(O_{r,c} - E_{r,c})^2}{E_{r,c}} \quad \chi^2 = 3.425$$

The degrees of freedom associated with this test statistic are:

$$df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1) \quad df = 1$$

The P-value is then calculated...

$$Pvalue = 1 - \text{chisq}(df, \chi^2) \quad Pvalue = 0.064 \quad \alpha = .1$$

**(9) Use the Decision Rule to state a Knowledge Claim**

$Pvalue > \alpha = 0$  indicates that  $Pvalue < \alpha$ , therefore reject the null hypothesis.

**(10) Consider the Practical Significance and State a Value Claim**

At a significance level of .1, the statistical analysis of the contingency table indicates that it is possible to reject the claim that homogeneity exists across the data set.

**(11) Use the Knowledge Claim and the Value Claim to Answer the Focus Question**

The facility's status of reporting (Report/Did Not Report) does appear to be related to whether the database used was the Standard Base Supply System versus some other system.

### **(1) State Focus Question to be Answered**

Is the facility's status of reporting (Report/Did Not Report) related to whether the database used was the database maintained by Bio-Environmental Engineering versus some other system?

### **(2) State Assumptions**

Assumptions:

1. Each observation is classified into exactly one cell.
2. The observations are observations on a random sample. Each observation has the same probability of being classified into cell (r,c) as any other observation.
3. The row totals are given, not random.

### **(3) State Hypothesis**

$H_0$ : The populations of Report/Did Not Report are homogeneous with respect to the database primarily used (BEE vs. Non-BEE)

$H_a$ : The populations of Report/Did Not Report are not homogeneous with respect to the database primarily used (BEE vs. Non-BEE)

### **(4) Specify a Test Statistic**

$$\chi^2 = \sum_r \sum_c \frac{(O_{r,c} - E_{r,c})^2}{E_{r,c}}$$

where  $r$  and  $c$  run over the rows and columns of the observed and expected value matrices.

The P-value will be computed as

$$Pvalue = 1 - \text{chisq}(df, \chi^2)$$

with  $df$  defined as  $df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1)$

### **(5) State the Distribution of the Test Statistic**

Under the null hypothesis, for a reasonable sample size, the statistic  $\chi^2$  computed above will approximately follow a  $\chi^2$  distribution having degrees of freedom

$$df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1)$$

### **(6) Formulate a Decision Rule**

If  $Pvalue > \alpha$  then do not reject  $H_0$  in favor of  $H_a$  at significance level  $\alpha$

**(7) Obtain a Random Sample**

BEE Non-BEE

$$BEE = \begin{pmatrix} 13 & 13 \\ 4 & 6 \end{pmatrix} \begin{matrix} \text{Reported} \\ \text{Did Not Report} \end{matrix}$$

**(8) Compute the Observed Value of the Test Statistic**

$$O = BEE$$

$$r = 0 \dots \text{rows}(O) - 1 \quad c = 0 \dots \text{cols}(O) - 1$$

$$N = \sum_r \sum_c BEE_{r,c} \quad N = 36$$

$$PC_c = \frac{\sum O^{<c>}}{N} \quad PC^T = (0.472 \quad 0.528) \quad \sum PC = 1$$

$$PR_r = \frac{\sum (O^T)^{<r>}}{N} \quad PR = \begin{pmatrix} 0.722 \\ 0.278 \end{pmatrix} \quad \sum PR = 1$$

Under the assumption of independence between column and row categories, the probability of being classified into the (r,c) cell will be

$$PR_r \cdot PC_c$$

the product of the corresponding row and column probabilities. A matrix of these probabilities can be built as follows:

$$P_{r,c} = PR_r \cdot PC_c \quad P = \begin{pmatrix} 0.341 & 0.381 \\ 0.131 & 0.147 \end{pmatrix} \quad \sum_r \sum_c P_{r,c} = 1$$

The expected cell values are then given by  $E = N \cdot P$

$$E = \begin{pmatrix} 12.278 & 13.722 \\ 4.722 & 5.278 \end{pmatrix}$$

and the test statistic can then easily be computed

$$\chi^2 = \sum_r \sum_c \frac{(O_{r,c} - E_{r,c})^2}{E_{r,c}} \quad \chi^2 = 0.29$$

The degrees of freedom associated with this test statistic are:

$$df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1) \quad df = 1$$

The P-value is then calculated...

$$Pvalue = 1 - \text{chisq}(df, \chi^2) \quad Pvalue = 0.59 \quad \alpha = .1$$

**(9) Use the Decision Rule to state a Knowledge Claim**

$Pvalue > \alpha = 1$  indicates that  $Pvalue > \alpha$ , therefore do not reject the null hypothesis.

**(10) Consider the Practical Significance and State a Value Claim**

At a significance level of .1, the statistical analysis of the contingency table indicates that it is not possible to reject the claim that homogeneity exists across the data set.

**(11) Use the Knowledge Claim and the Value Claim to Answer the Focus Question**

The facility's status of reporting (Report/Did Not Report) does not appear to be related to whether the database used was the database maintained by Bio-Environmental Engineering versus some other system.



**(1) State Focus Question to be Answered**

Is the facility's status of reporting (Report/Did Not Report) related to whether the TRI POC thought the data from the primary database used was complete or had deficiencies?

**(2) State Assumptions**

Assumptions:

1. Each observation is classified into exactly one cell.
2. The observations are observations on a random sample. Each observation has the same probability of being classified into cell (r,c) as any other observation.
3. The row totals are given, not random.

**(3) State Hypothesis**

**H<sub>0</sub>:** The populations of Report/Did Not Report are homogeneous with respect to whether the TRI POC thought the data from the primary database used was complete or had deficiencies

**H<sub>a</sub>:** The populations of Report/Did Not Report are not homogeneous with respect to whether the TRI POC thought the data from the primary database used was complete or had deficiencies

**(4) Specify a Test Statistic**

$$\chi^2 = \sum_r \sum_c \frac{(O_{r,c} - E_{r,c})^2}{E_{r,c}}$$

where **r** and **c** run over the rows and columns of the observed and expected value matrices.

The P-value will be computed as

$$Pvalue = 1 - \text{chisq}(df, \chi^2)$$

with df defined as  $df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1)$

**(5) State the Distribution of the Test Statistic**

Under the null hypothesis, for a reasonable sample size, the statistic  $\chi^2$  computed above will approximately follow a  $\chi^2$  distribution having degrees of freedom

$$df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1)$$

**(6) Formulate a Decision Rule**

If  $Pvalue > \alpha$  then do not reject  $H_0$  in favor of  $H_a$  at significance level  $\alpha$

**(7) Obtain a Random Sample**

Complete Non-Complete

$$\text{Complete} = \begin{pmatrix} 8 & 18 \\ 0 & 10 \end{pmatrix} \begin{matrix} \text{Reported} \\ \text{Did Not Report} \end{matrix}$$

**(8) Compute the Observed Value of the Test Statistic**

$O = \text{Complete}$

$r = 0 \dots \text{rows}(O) - 1$   $c = 0 \dots \text{cols}(O) - 1$

$$N = \sum_r \sum_c \text{Complete}_{r,c} \quad N = 36$$

$$PC_c = \frac{\sum O^{<c>}}{N} \quad PC^T = (0.222 \quad 0.778) \quad \sum PC = 1$$

$$PR_r = \frac{\sum (O^T)^{<r>}}{N} \quad PR = \begin{pmatrix} 0.722 \\ 0.278 \end{pmatrix} \quad \sum PR = 1$$

Under the assumption of independence between column and row categories, the probability of being classified into the  $(r,c)$  cell will be

$$PR_r \cdot PC_c$$

the product of the corresponding row and column probabilities. A matrix of these probabilities can be built as follows:

$$P_{r,c} = PR_r \cdot PC_c \quad P = \begin{pmatrix} 0.16 & 0.562 \\ 0.062 & 0.216 \end{pmatrix} \quad \sum_r \sum_c P_{r,c} = 1$$

The expected cell values are then given by  $E = N \cdot P$

$$E = \begin{pmatrix} 5.778 & 20.222 \\ 2.222 & 7.778 \end{pmatrix}$$

and the test statistic can then easily be computed

$$\chi^2 = \sum_r \sum_c \frac{(O_{r,c} - E_{r,c})^2}{E_{r,c}} \quad \chi^2 = 3.956$$

The degrees of freedom associated with this test statistic are:

$$df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1) \quad df = 1$$

The P-value is then calculated...

$$Pvalue = 1 - \text{chisq}(df, \chi^2) \quad Pvalue = 0.047 \quad \alpha = .1$$

**(9) Use the Decision Rule to state a Knowledge Claim**

$Pvalue > \alpha = 0$  indicates that  $Pvalue < \alpha$ , therefore reject the null hypothesis.

**(10) Consider the Practical Significance and State a Value Claim**

At a significance level of .1, the statistical analysis of the contingency table indicates that it is possible to reject the claim that homogeneity exists across the data set.

**(11) Use the Knowledge Claim and the Value Claim to Answer the Focus Question**

The facility's status of reporting (Report/Did Not Report) does appear to be related to whether the TRI POC thought the data from the primary database used was complete or had deficiencies.

**(1) State Focus Question to be Answered**

Is the facility's status of reporting (Report/Did Not Report) related to which guidance document the TRI POCs considered the most important or helpful?

**(2) State Assumptions**

Assumptions:

1. Each observation is classified into exactly one cell.
2. The observations are observations on a random sample. Each observation has the same probability of being classified into cell (r,c) as any other observation.
3. The row totals are given, not random.

**(3) State Hypothesis**

$H_0$ : The populations of Report/Did Not Report are homogeneous with respect to which guidance document the TRI POCs considered the most important or helpful

$H_a$ : The populations of Report/Did Not Report are not homogeneous with respect to which guidance document the TRI POCs considered the most important or helpful

**(4) Specify a Test Statistic**

$$\chi^2 = \sum_r \sum_c \frac{(O_{r,c} - E_{r,c})^2}{E_{r,c}}$$

where  $r$  and  $c$  run over the rows and columns of the observed and expected value matrices.

The P-value will be computed as

$$Pvalue = 1 - \text{chisq}(df, \chi^2)$$

with  $df$  defined as  $df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1)$

**(5) State the Distribution of the Test Statistic**

Under the null hypothesis, for a reasonable sample size, the statistic  $\chi^2$  computed above will approximately follow a  $\chi^2$  distribution having degrees of freedom

$$df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1)$$

**(6) Formulate a Decision Rule**

If  $Pvalue > \alpha$  then do not reject  $H_0$  in favor of  $H_a$  at significance level  $\alpha$

**(7) Obtain a Random Sample**

$$\text{Guidance} = \begin{pmatrix} 11 & 1 & 4 & 1 & 2 & 6 & 1 \\ 5 & 0 & 1 & 2 & 0 & 1 & 1 \end{pmatrix} \begin{matrix} \text{Reported} \\ \text{Did Not Report} \end{matrix}$$

The columns from left to right represent AF, DoD, GAIA Handbook, GAIA Manual, EPA, Form R instructions and the State of Florida guidance.

**(8) Compute the Observed Value of the Test Statistic**

$O = \text{Guidance}$

$r = 0.. \text{rows}(O) - 1 \quad c = 0.. \text{cols}(O) - 1$

$$N = \sum_r \sum_c \text{Guidance}_{r,c} \quad N = 36$$

$$PC_c = \frac{\sum O^{<c>}}{N} \quad PC^T = (0.444 \quad 0.028 \quad 0.139 \quad 0.083 \quad 0.056 \quad 0.194 \quad 0.056)$$

$$\sum PC = 1$$

$$PR_r = \frac{\sum (O^T)^{<r>}}{N} \quad PR = \begin{pmatrix} 0.722 \\ 0.278 \end{pmatrix} \quad \sum PR = 1$$

Under the assumption of independence between column and row categories, the probability of being classified into the  $(r,c)$  cell will be

$$PR_r \cdot PC_c$$

the product of the corresponding row and column probabilities. A matrix of these probabilities can then be built...

$$P_{r,c} = PR_r \cdot PC_c \quad P = \begin{pmatrix} 0.321 & 0.02 & 0.1 & 0.06 & 0.04 & 0.14 & 0.04 \\ 0.123 & 0.008 & 0.039 & 0.023 & 0.015 & 0.054 & 0.015 \end{pmatrix}$$

$$\sum_r \sum_c P_{r,c} = 1$$

The expected cell values are then given by  $E = N \cdot P$

$$E = \begin{pmatrix} 11.556 & 0.722 & 3.611 & 2.167 & 1.444 & 5.056 & 1.444 \\ 4.444 & 0.278 & 1.389 & 0.833 & 0.556 & 1.944 & 0.556 \end{pmatrix}$$

and the test statistic can then easily be computed

$$\chi^2 = \sum_r \sum_c \frac{(O_{r,c} - E_{r,c})^2}{E_{r,c}} \quad \chi^2 = 4.79$$

The degrees of freedom associated with this test statistic are:

$$df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1) \quad df = 6$$

The P-value is then calculated...

$$Pvalue = 1 - \text{chisq}(df, \chi^2) \quad Pvalue = 0.571 \quad \alpha = .1$$

**(9) Use the Decision Rule to state a Knowledge Claim**

$Pvalue > \alpha = 1$  indicates that  $Pvalue > \alpha$ , therefore do not reject the null hypothesis.

**(10) Consider the Practical Significance and State a Value Claim**

At a significance level of .1, the statistical analysis of the contingency table indicates that it is not possible to reject the claim that homogeneity exists across the data set.

**(11) Use the Knowledge Claim and the Value Claim to Answer the Focus Question**

The facility's status of reporting (Report/Did Not Report) does not appear to be related to which guidance document the TRI POCs considered the most important or helpful.

**(1) State Focus Question to be Answered**

Is the facility's status of reporting (Report/Did Not Report) related to whether the guidance used was developed by the Air Force or some other source?

**(2) State Assumptions**

Assumptions:

1. Each observation is classified into exactly one cell.
2. The observations are observations on a random sample. Each observation has the same probability of being classified into cell (r,c) as any other observation.
3. The row totals are given, not random.

**(3) State Hypothesis**

$H_0$ : The populations of Report/Did Not Report are homogeneous with respect to whether the guidance used was developed by the Air Force or some other source

$H_a$ : The populations of Report/Did Not Report are not homogeneous with respect to whether the guidance used was developed by the Air Force or some other source

**(4) Specify a Test Statistic**

$$\chi^2 = \sum_r \sum_c \frac{(O_{r,c} - E_{r,c})^2}{E_{r,c}}$$

where  $r$  and  $c$  run over the rows and columns of the observed and expected value matrices.

The P-value will be computed as

$$Pvalue = 1 - \text{chisq}(df, \chi^2)$$

with  $df$  defined as  $df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1)$

**(5) State the Distribution of the Test Statistic**

Under the null hypothesis, for a reasonable sample size, the statistic  $\chi^2$  computed above will approximately follow a  $\chi^2$  distribution having degrees of freedom

$$df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1)$$

**(6) Formulate a Decision Rule**

If  $Pvalue > \alpha$  then do not reject  $H_0$  in favor of  $H_a$  at significance level  $\alpha$

**(7) Obtain a Random Sample**

AF Non-AF

$$AF = \begin{pmatrix} 11 & 15 \\ 5 & 5 \end{pmatrix} \begin{matrix} \text{Reported} \\ \text{Did Not Report} \end{matrix}$$

**(8) Compute the Observed Value of the Test Statistic**

$$O = AF$$

$$r = 0 \dots \text{rows}(O) - 1 \quad c = 0 \dots \text{cols}(O) - 1$$

$$N = \sum_r \sum_c AF_{r,c} \quad N = 36$$

$$PC_c = \frac{\sum O^{<c>}}{N} \quad PC^T = (0.444 \quad 0.556) \quad \sum PC = 1$$

$$PR_r = \frac{\sum (O^T)^{<r>}}{N} \quad PR = \begin{pmatrix} 0.722 \\ 0.278 \end{pmatrix} \quad \sum PR = 1$$

Under the assumption of independence between column and row categories, the probability of being classified into the (r,c) cell will be

$$PR_r \cdot PC_c$$

the product of the corresponding row and column probabilities. A matrix of these probabilities can be built as follows:

$$P_{r,c} := PR_r \cdot PC_c \quad P = \begin{pmatrix} 0.321 & 0.401 \\ 0.123 & 0.154 \end{pmatrix} \quad \sum_r \sum_c P_{r,c} = 1$$

The expected cell values are then given by  $E := N \cdot P$

$$E = \begin{pmatrix} 11.556 & 14.444 \\ 4.444 & 5.556 \end{pmatrix}$$



and the test statistic can then easily be computed

$$\chi^2 = \sum_r \sum_c \frac{(O_{r,c} - E_{r,c})^2}{E_{r,c}} \quad \chi^2 = 0.173$$

The degrees of freedom associated with this test statistic are:

$$df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1) \quad df = 1$$

The P-value is then calculated...

$$Pvalue = 1 - \text{chisq}(df, \chi^2) \quad Pvalue = 0.677 \quad \alpha = .1$$

**(9) Use the Decision Rule to state a Knowledge Claim**

$Pvalue > \alpha = 1$  indicates that  $Pvalue > \alpha$ , therefore do not reject the null hypothesis.

**(10) Consider the Practical Significance and State a Value Claim**

At a significance level of .1, the statistical analysis of the contingency table indicates that it is not possible to reject the claim that homogeneity exists across the data set.

**(11) Use the Knowledge Claim and the Value Claim to Answer the Focus Question**

The facility's status of reporting (Report/Did Not Report) does not appear to be related to whether the guidance used was developed by the Air Force or some other source

### **(1) State Focus Question to be Answered**

Is the facility's status of reporting (Report/Did Not Report) related to whether the guidance used was developed by the military (either the DoD or Air Force) or some other source?

### **(2) State Assumptions**

Assumptions:

1. Each observation is classified into exactly one cell.
2. The observations are observations on a random sample. Each observation has the same probability of being classified into cell (r,c) as any other observation.
3. The row totals are given, not random.

### **(3) State Hypothesis**

**H<sub>0</sub>:** The populations of Report/Did Not Report are homogeneous with respect to whether the guidance used was developed by the military (either the DoD or Air Force) or some other source

**H<sub>a</sub>:** The populations of Report/Did Not Report are not homogeneous with respect to whether the guidance used was developed by the military (either the DoD or Air Force) or some other source

### **(4) Specify a Test Statistic**

$$\chi^2 = \sum_r \sum_c \frac{(O_{r,c} - E_{r,c})^2}{E_{r,c}}$$

where *r* and *c* run over the rows and columns of the observed and expected value matrices.

The P-value will be computed as

$$Pvalue = 1 - \text{chisq}(df, \chi^2)$$

with df defined as  $df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1)$

### **(5) State the Distribution of the Test Statistic**

Under the null hypothesis, for a reasonable sample size, the statistic  $\chi^2$  computed above will approximately follow a  $\chi^2$  distribution having degrees of freedom

$$df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1)$$

### **(6) Formulate a Decision Rule**

If  $Pvalue > \alpha$  then do not reject  $H_0$  in favor of  $H_a$  at significance level  $\alpha$

**(7) Obtain a Random Sample**

$$\text{Military} = \begin{matrix} & \begin{matrix} \text{Military} & \text{Non-Military} \end{matrix} \\ \begin{matrix} \text{Military} \\ \text{Non-Military} \end{matrix} & \begin{pmatrix} 12 & 14 \\ 5 & 5 \end{pmatrix} \end{matrix} \begin{matrix} \text{Reported} \\ \text{Did Not Report} \end{matrix}$$

**(8) Compute the Observed Value of the Test Statistic**

$O = \text{Military}$

$r = 0 \dots \text{rows}(O) - 1$   $c = 0 \dots \text{cols}(O) - 1$

$$N = \sum_r \sum_c \text{Military}_{r,c} \quad N = 36$$

$$PC_c = \frac{\sum_r O_{r,c}}{N} \quad PC^T = (0.472 \quad 0.528) \quad \sum PC = 1$$

$$PR_r = \frac{\sum_c (O^T)_{r,c}}{N} \quad PR = \begin{pmatrix} 0.722 \\ 0.278 \end{pmatrix} \quad \sum PR = 1$$

Under the assumption of independence between column and row categories, the probability of being classified into the  $(r,c)$  cell will be

$$PR_r \cdot PC_c$$

the product of the corresponding row and column probabilities. A matrix of these probabilities can be built as follows:

$$P_{r,c} = PR_r \cdot PC_c \quad P = \begin{pmatrix} 0.341 & 0.381 \\ 0.131 & 0.147 \end{pmatrix} \quad \sum_r \sum_c P_{r,c} = 1$$

The expected cell values are then given by  $E = N \cdot P$

$$E = \begin{pmatrix} 12.278 & 13.722 \\ 4.722 & 5.278 \end{pmatrix}$$

and the test statistic can then easily be computed

$$\chi^2 = \sum_r \sum_c \frac{(O_{r,c} - E_{r,c})^2}{E_{r,c}} \quad \chi^2 = 0.043$$

The degrees of freedom associated with this test statistic are:

$$df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1) \quad df = 1$$

The P-value is then calculated...

$$Pvalue = 1 - \text{chisq}(df, \chi^2) \quad Pvalue = 0.836 \quad \alpha = .1$$

#### **(9) Use the Decision Rule to state a Knowledge Claim**

$Pvalue > \alpha = 1$  indicates that  $Pvalue > \alpha$ , therefore do not reject the null hypothesis.

#### **(10) Consider the Practical Significance and State a Value Claim**

At a significance level of .1, the statistical analysis of the contingency table indicates that it is not possible to reject the claim that homogeneity exists across the data set.

#### **(11) Use the Knowledge Claim and the Value Claim to Answer the Focus Question**

The facility's status of reporting (Report/Did Not Report) does not appear to be related to whether the guidance used was developed by the military (either the DoD or Air Force) or some other source.

### **(1) State Focus Question to be Answered**

Is the facility's status of reporting (Report/Did Not Report) related to whether the guidance used was developed by the EPA or some other source?

### **(2) State Assumptions**

Assumptions:

1. Each observation is classified into exactly one cell.
2. The observations are observations on a random sample. Each observation has the same probability of being classified into cell (r,c) as any other observation.
3. The row totals are given, not random.

### **(3) State Hypothesis**

$H_0$ : The populations of Report/Did Not Report are homogeneous with respect to whether the guidance used was developed by the EPA or some other source

$H_a$ : The populations of Report/Did Not Report are not homogeneous with respect to whether the guidance used was developed by the EPA or some other source

### **(4) Specify a Test Statistic**

$$\chi^2 = \sum_r \sum_c \frac{(O_{r,c} - E_{r,c})^2}{E_{r,c}}$$

where  $r$  and  $c$  run over the rows and columns of the observed and expected value matrices.

The P-value will be computed as

$$Pvalue = 1 - \text{chisq}(df, \chi^2)$$

with  $df$  defined as  $df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1)$

### **(5) State the Distribution of the Test Statistic**

Under the null hypothesis, for a reasonable sample size, the statistic  $\chi^2$  computed above will approximately follow a  $\chi^2$  distribution having degrees of freedom

$$df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1)$$

### **(6) Formulate a Decision Rule**

If  $Pvalue > \alpha$  then do not reject  $H_0$  in favor of  $H_a$  at significance level  $\alpha$

**(7) Obtain a Random Sample**

$$\begin{array}{cc} \text{EPA} & \text{Non-EPA} \\ \text{EPA} = \begin{pmatrix} 8 & 18 \\ 1 & 9 \end{pmatrix} & \begin{array}{l} \text{Reported} \\ \text{Did Not Report} \end{array} \end{array}$$

**(8) Compute the Observed Value of the Test Statistic**

$$O = \text{EPA}$$

$$r = 0 \dots \text{rows}(O) - 1 \quad c = 0 \dots \text{cols}(O) - 1$$

$$N = \sum_r \sum_c \text{EPA}_{r,c} \quad N = 36$$

$$\text{PC}_c = \frac{\sum O^{<c>}}{N} \quad \text{PC}^T = (0.25 \quad 0.75) \quad \sum \text{PC} = 1$$

$$\text{PR}_r = \frac{\sum (O^T)^{<r>}}{N} \quad \text{PR} = \begin{pmatrix} 0.722 \\ 0.278 \end{pmatrix} \quad \sum \text{PR} = 1$$

Under the assumption of independence between column and row categories, the probability of being classified into the (r,c) cell will be

$$\text{PR}_r \cdot \text{PC}_c$$

the product of the corresponding row and column probabilities. A matrix of these probabilities can be built as follows:

$$P_{r,c} = \text{PR}_r \cdot \text{PC}_c \quad P = \begin{pmatrix} 0.181 & 0.542 \\ 0.069 & 0.208 \end{pmatrix} \quad \sum_r \sum_c P_{r,c} = 1$$

The expected cell values are then given by  $E = N \cdot P$

$$E = \begin{pmatrix} 6.5 & 19.5 \\ 2.5 & 7.5 \end{pmatrix}$$

and the test statistic can then easily be computed

$$\chi^2 = \sum_r \sum_c \frac{(O_{r,c} - E_{r,c})^2}{E_{r,c}} \quad \chi^2 = 1.662$$

The degrees of freedom associated with this test statistic are:

$$df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1) \quad df = 1$$

The P-value is then calculated...

$$Pvalue = 1 - \text{chisq}(df, \chi^2) \quad Pvalue = 0.197 \quad \alpha = .1$$

**(9) Use the Decision Rule to state a Knowledge Claim**

$Pvalue > \alpha = 1$  indicates that  $Pvalue > \alpha$ , therefore do not reject the null hypothesis.

**(10) Consider the Practical Significance and State a Value Claim**

At a significance level of .1, the statistical analysis of the contingency table indicates that it is not possible to reject the claim that homogeneity exists across the data set.

**(11) Use the Knowledge Claim and the Value Claim to Answer the Focus Question**

The facility's status of reporting (Report/Did Not Report) does not appear to be related to whether the guidance used was developed by the EPA or some other source.

**(1) State Focus Question to be Answered**

Is the facility's status of reporting (Report/Did Not Report) related to whether the guidance used was developed by a federal agency or some other source?

**(2) State Assumptions**

Assumptions:

1. Each observation is classified into exactly one cell.
2. The observations are observations on a random sample. Each observation has the same probability of being classified into cell (r,c) as any other observation.
3. The row totals are given, not random.

**(3) State Hypothesis**

$H_0$ : The populations of Report/Did Not Report are homogeneous with respect to whether the guidance used was developed by a federal agency or some other source

$H_a$ : The populations of Report/Did Not Report are not homogeneous with respect to whether the guidance used was developed by a federal agency or some other source

**(4) Specify a Test Statistic**

$$\chi^2 = \sum_r \sum_c \frac{(O_{r,c} - E_{r,c})^2}{E_{r,c}}$$

where  $r$  and  $c$  run over the rows and columns of the observed and expected value matrices.

The P-value will be computed as

$$\text{Pvalue} = 1 - \text{chisq}(\text{df}, \chi^2)$$

with df defined as  $\text{df} = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1)$

**(5) State the Distribution of the Test Statistic**

Under the null hypothesis, for a reasonable sample size, the statistic  $\chi^2$  computed above will approximately follow a  $\chi^2$  distribution having degrees of freedom

$$\text{df} = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1)$$

**(6) Formulate a Decision Rule**

If  $\text{Pvalue} > \alpha$  then do not reject  $H_0$  in favor of  $H_a$  at significance level  $\alpha$



**(7) Obtain a Random Sample**

$$\text{Federal} = \begin{matrix} & \begin{matrix} \text{Federal} & \text{Non-Federal} \end{matrix} \\ \begin{matrix} \text{Federal} \\ \text{Non-Federal} \end{matrix} & \begin{pmatrix} 20 & 6 \\ 6 & 4 \end{pmatrix} \end{matrix} \begin{matrix} \text{Reported} \\ \text{Did Not Report} \end{matrix}$$

**(8) Compute the Observed Value of the Test Statistic**

$$O := \text{Federal}$$

$$r := 0.. \text{rows}(O) - 1 \quad c := 0.. \text{cols}(O) - 1$$

$$N := \sum_r \sum_c \text{Federal}_{r,c} \quad N = 36$$

$$PC_c := \frac{\sum O^{<c>}}{N} \quad PC^T = \begin{pmatrix} 0.722 & 0.278 \end{pmatrix} \quad \sum PC = 1$$

$$PR_r := \frac{\sum (O^T)^{<r>}}{N} \quad PR = \begin{pmatrix} 0.722 \\ 0.278 \end{pmatrix} \quad \sum PR = 1$$

Under the assumption of independence between column and row categories, the probability of being classified into the (r,c) cell will be

$$PR_r \cdot PC_c$$

the product of the corresponding row and column probabilities. A matrix of these probabilities can be built as follows:

$$P_{r,c} := PR_r \cdot PC_c \quad P = \begin{pmatrix} 0.522 & 0.201 \\ 0.201 & 0.077 \end{pmatrix} \quad \sum_r \sum_c P_{r,c} = 1$$

The expected cell values are then given by  $E = N \cdot P$

$$E = \begin{pmatrix} 18.778 & 7.222 \\ 7.222 & 2.778 \end{pmatrix}$$

and the test statistic can then easily be computed

$$\chi^2 = \sum_r \sum_c \frac{(O_{r,c} - E_{r,c})^2}{E_{r,c}} \quad \chi^2 = 1.031$$

The degrees of freedom associated with this test statistic are:

$$df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1) \quad df = 1$$

The P-value is then calculated...

$$Pvalue = 1 - \text{chisq}(df, \chi^2) \quad Pvalue = 0.31 \quad \alpha = .1$$

**(9) Use the Decision Rule to state a Knowledge Claim**

$Pvalue > \alpha = 1$  indicates that  $Pvalue > \alpha$ , therefore do not reject the null hypothesis.

**(10) Consider the Practical Significance and State a Value Claim**

At a significance level of .1, the statistical analysis of the contingency table indicates that it is not possible to reject the claim that homogeneity exists across the data set.

**(11) Use the Knowledge Claim and the Value Claim to Answer the Focus Question**

The facility's status of reporting (Report/Did Not Report) does not appear to be related to whether the guidance used was developed by a federal agency or some other source.

**(1) State Focus Question to be Answered**

Is the facility's status of reporting (Report/Did Not Report) related to whether the guidance used was developed by GAIA Corp. or some other source?

**(2) State Assumptions**

Assumptions:

1. Each observation is classified into exactly one cell.
2. The observations are observations on a random sample. Each observation has the same probability of being classified into cell (r,c) as any other observation.
3. The row totals are given, not random.

**(3) State Hypothesis**

$H_0$ : The populations of Report/Did Not Report are homogeneous with respect to whether the guidance used was developed by GAIA Corp. or some other source

$H_a$ : The populations of Report/Did Not Report are not homogeneous with respect to whether the guidance used was developed by the GAIA Corp. or some other source

**(4) Specify a Test Statistic**

$$\chi^2 = \sum_r \sum_c \frac{(O_{r,c} - E_{r,c})^2}{E_{r,c}}$$

where  $r$  and  $c$  run over the rows and columns of the observed and expected value matrices.

The P-value will be computed as

$$Pvalue = 1 - \text{chisq}(df, \chi^2)$$

with  $df$  defined as  $df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1)$

**(5) State the Distribution of the Test Statistic**

Under the null hypothesis, for a reasonable sample size, the statistic  $\chi^2$  computed above will approximately follow a  $\chi^2$  distribution having degrees of freedom

$$df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1)$$

**(6) Formulate a Decision Rule**

If  $Pvalue > \alpha$  then do not reject  $H_0$  in favor of  $H_a$  at significance level  $\alpha$

**(7) Obtain a Random Sample**

$$\text{GAIA} = \begin{pmatrix} 5 & 21 \\ 3 & 7 \end{pmatrix} \begin{matrix} \text{Reported} \\ \text{Did Not Report} \end{matrix}$$

**(8) Compute the Observed Value of the Test Statistic**

$$O := \text{GAIA}$$

$$r := 0 \dots \text{rows}(O) - 1 \quad c := 0 \dots \text{cols}(O) - 1$$

$$N := \sum_r \sum_c \text{GAIA}_{r,c} \quad N = 36$$

$$\text{PC}_c = \frac{\sum O^{<c>}}{N} \quad \text{PC}^T = (0.222 \quad 0.778) \quad \sum \text{PC} = 1$$

$$\text{PR}_r = \frac{\sum (O^T)^{<r>}}{N} \quad \text{PR} = \begin{pmatrix} 0.722 \\ 0.278 \end{pmatrix} \quad \sum \text{PR} = 1$$

Under the assumption of independence between column and row categories, the probability of being classified into the (r,c) cell will be

$$\text{PR}_r \cdot \text{PC}_c$$

the product of the corresponding row and column probabilities. A matrix of these probabilities can be built as follows:

$$P_{r,c} := \text{PR}_r \cdot \text{PC}_c \quad P = \begin{pmatrix} 0.16 & 0.562 \\ 0.062 & 0.216 \end{pmatrix} \quad \sum_r \sum_c P_{r,c} = 1$$

The expected cell values are then given by  $E := N \cdot P$

$$E = \begin{pmatrix} 5.778 & 20.222 \\ 2.222 & 7.778 \end{pmatrix}$$

and the test statistic can then easily be computed

$$\chi^2 = \sum_r \sum_c \frac{(O_{r,c} - E_{r,c})^2}{E_{r,c}} \quad \chi^2 = 0.485$$

The degrees of freedom associated with this test statistic are:

$$df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1) \quad df = 1$$

The P-value is then calculated...

$$P\text{value} = 1 - \text{chisq}(df, \chi^2) \quad P\text{value} = 0.486 \quad \alpha = .1$$

**(9) Use the Decision Rule to state a Knowledge Claim**

$P\text{value} > \alpha = 1$  indicates that  $P\text{value} > \alpha$ , therefore do not reject the null hypothesis.

**(10) Consider the Practical Significance and State a Value Claim**

At a significance level of .1, the statistical analysis of the contingency table indicates that it is not possible to reject the claim that homogeneity exists across the data set.

**(11) Use the Knowledge Claim and the Value Claim to Answer the Focus Question**

The facility's status of reporting (Report/Did Not Report) does not appear to be related to whether the guidance used was developed by the GAIA Corp or some other source.

**(1) State Focus Question to be Answered**

Is the facility's status of reporting (Report/Did Not Report) related to whether the TRI POC developed a list of processes on the facility that generate TRI releases?

**(2) State Assumptions**

Assumptions:

1. Each observation is classified into exactly one cell.
2. The observations are observations on a random sample. Each observation has the same probability of being classified into cell (r,c) as any other observation.
3. The row totals are given, not random.

**(3) State Hypothesis**

$H_0$ : The populations of Report/Did Not Report are homogeneous with respect to whether the TRI POC developed a list of processes on the facility that generate TRI releases

$H_a$ : The populations of Report/Did Not Report are not homogeneous with respect to whether the TRI POC developed a list of processes on the facility that generate TRI releases

**(4) Specify a Test Statistic**

$$\chi^2 = \sum_r \sum_c \frac{(O_{r,c} - E_{r,c})^2}{E_{r,c}}$$

where  $r$  and  $c$  run over the rows and columns of the observed and expected value matrices.

The P-value will be computed as

$$Pvalue = 1 - \text{chisq}(df, \chi^2)$$

with  $df$  defined as  $df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1)$

**(5) State the Distribution of the Test Statistic**

Under the null hypothesis, for a reasonable sample size, the statistic  $\chi^2$  computed above will approximately follow a  $\chi^2$  distribution having degrees of freedom

$$df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1)$$

**(6) Formulate a Decision Rule**

If  $Pvalue > \alpha$  then do not reject  $H_0$  in favor of  $H_a$  at significance level  $\alpha$

**(7) Obtain a Random Sample**

$$\text{List} := \begin{matrix} \text{Yes} & \text{No} \\ \left( \begin{matrix} 16 & 10 \\ 5 & 5 \end{matrix} \right) & \begin{matrix} \text{Reported} \\ \text{Did Not Report} \end{matrix} \end{matrix}$$

**(8) Compute the Observed Value of the Test Statistic**

$$O := \text{List}$$

$$r := 0 \dots \text{rows}(O) - 1 \quad c := 0 \dots \text{cols}(O) - 1$$

$$N := \sum_r \sum_c \text{List}_{r,c} \quad N = 36$$

$$\text{PC}_c := \frac{\sum O^{<c>}}{N} \quad \text{PC}^T = \begin{pmatrix} 0.583 & 0.417 \end{pmatrix} \quad \sum \text{PC} = 1$$

$$\text{PR}_r := \frac{\sum (O^T)^{<r>}}{N} \quad \text{PR} = \begin{pmatrix} 0.722 \\ 0.278 \end{pmatrix} \quad \sum \text{PR} = 1$$

Under the assumption of independence between column and row categories, the probability of being classified into the (r,c) cell will be

$$\text{PR}_r \cdot \text{PC}_c$$

the product of the corresponding row and column probabilities. A matrix of these probabilities can be built as follows:

$$P_{r,c} := \text{PR}_r \cdot \text{PC}_c \quad P = \begin{pmatrix} 0.421 & 0.301 \\ 0.162 & 0.116 \end{pmatrix} \quad \sum_r \sum_c P_{r,c} = 1$$

The expected cell values are then given by  $E := N \cdot P$

$$E = \begin{pmatrix} 15.167 & 10.833 \\ 5.833 & 4.167 \end{pmatrix}$$

and the test statistic can then easily be computed

$$\chi^2 = \sum_r \sum_c \frac{(O_{r,c} - E_{r,c})^2}{E_{r,c}} \quad \chi^2 = 0.396$$

The degrees of freedom associated with this test statistic are:

$$df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1) \quad df = 1$$

The P-value is then calculated...

$$P\text{value} = 1 - \text{chisq}(df, \chi^2) \quad P\text{value} = 0.529 \quad \alpha = .1$$

**(9) Use the Decision Rule to state a Knowledge Claim**

$P\text{value} > \alpha = 1$  indicates that  $P\text{value} > \alpha$ , therefore do not reject the null hypothesis.

**(10) Consider the Practical Significance and State a Value Claim**

At a significance level of .1, the statistical analysis of the contingency table indicates that it is not possible to reject the claim that homogeneity exists across the data set.

**(11) Use the Knowledge Claim and the Value Claim to Answer the Focus Question**

The facility's status of reporting (Report/Did Not Report) does not appear to be related to whether the TRI POC developed a list of processes on the facility that generate TRI releases.



### **(1) State Focus Question to be Answered**

Is the facility's status of reporting (Report/Did Not Report) related to whether the TRI POC visited the sites of chemical releases on the facility during the TRI process?

### **(2) State Assumptions**

Assumptions:

1. Each observation is classified into exactly one cell.
2. The observations are observations on a random sample. Each observation has the same probability of being classified into cell (r,c) as any other observation.
3. The row totals are given, not random.

### **(3) State Hypothesis**

**H<sub>0</sub>:** The populations of Report/Did Not Report are homogeneous with respect to whether the TRI POC visited the sites of chemical releases on the facility during the TRI process

**H<sub>a</sub>:** The populations of Report/Did Not Report are not homogeneous with respect to whether the TRI POC visited the sites of chemical releases on the facility during the TRI process.

### **(4) Specify a Test Statistic**

$$\chi^2 = \sum_r \sum_c \frac{(O_{r,c} - E_{r,c})^2}{E_{r,c}}$$

where *r* and *c* run over the rows and columns of the observed and expected value matrices.

The P-value will be computed as

$$\text{Pvalue} = 1 - \text{chisq}(\text{df}, \chi^2)$$

with df defined as  $\text{df} = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1)$

### **(5) State the Distribution of the Test Statistic**

Under the null hypothesis, for a reasonable sample size, the statistic  $\chi^2$  computed above will approximately follow a  $\chi^2$  distribution having degrees of freedom

$$\text{df} = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1)$$

### **(6) Formulate a Decision Rule**

If  $\text{Pvalue} > \alpha$  then do not reject  $H_0$  in favor of  $H_a$  at significance level  $\alpha$

**(7) Obtain a Random Sample**

$$\begin{array}{cc} \text{Yes} & \text{No} \\ \text{Visit} := \begin{pmatrix} 21 & 5 \\ 8 & 2 \end{pmatrix} & \begin{array}{l} \text{Reported} \\ \text{Did Not Report} \end{array} \end{array}$$

**(8) Compute the Observed Value of the Test Statistic**

$$O := \text{Visit}$$

$$r := 0 \dots \text{rows}(O) - 1 \quad c := 0 \dots \text{cols}(O) - 1$$

$$N := \sum_r \sum_c \text{Visit}_{r,c} \quad N = 36$$

$$PC_c := \frac{\sum O^{<c>}}{N} \quad PC^T = (0.806 \quad 0.194) \quad \sum PC = 1$$

$$PR_r := \frac{\sum (O^T)^{<r>}}{N} \quad PR = \begin{pmatrix} 0.722 \\ 0.278 \end{pmatrix} \quad \sum PR = 1$$

Under the assumption of independence between column and row categories, the probability of being classified into the (r,c) cell will be

$$PR_r \cdot PC_c$$

the product of the corresponding row and column probabilities. A matrix of these probabilities can be built as follows:

$$P_{r,c} := PR_r \cdot PC_c \quad P = \begin{pmatrix} 0.582 & 0.14 \\ 0.224 & 0.054 \end{pmatrix} \quad \sum_r \sum_c P_{r,c} = 1$$

The expected cell values are then given by  $E := N \cdot P$

$$E = \begin{pmatrix} 20.944 & 5.056 \\ 8.056 & 1.944 \end{pmatrix}$$

and the test statistic can then easily be computed

$$\chi^2 = \sum_r \sum_c \frac{(O_{r,c} - E_{r,c})^2}{E_{r,c}} \quad \chi^2 = 0.003$$

The degrees of freedom associated with this test statistic are:

$$df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1) \quad df = 1$$

The P-value is then calculated...

$$Pvalue = 1 - \text{chisq}(df, \chi^2) \quad Pvalue = 0.958 \quad \alpha = .1$$

**(9) Use the Decision Rule to state a Knowledge Claim**

$Pvalue > \alpha = 1$  indicates that  $Pvalue > \alpha$ , therefore do not reject the null hypothesis.

**(10) Consider the Practical Significance and State a Value Claim**

At a significance level of .1, the statistical analysis of the contingency table indicates that it is not possible to reject the claim that homogeneity exists across the data set.

**(11) Use the Knowledge Claim and the Value Claim to Answer the Focus Question**

The facility's status of reporting (Report/Did Not Report) does not appear to be related to whether the TRI POC visited the sites of chemical releases on the facility during the TRI process.

**(1) State Focus Question to be Answered**

Is the facility's status of reporting (Report/Did Not Report) related to which major command the reporting facility belongs?

**(2) State Assumptions**

Assumptions:

1. Each observation is classified into exactly one cell.
2. The observations are observations on a random sample. Each observation has the same probability of being classified into cell (r,c) as any other observation.
3. The row totals are given, not random.

**(3) State Hypothesis**

$H_0$ : The populations of Report/Did Not Report are homogeneous with respect to which major command the reporting facility belongs.

$H_a$ : The populations of Report/Did Not Report are not homogeneous with respect to which major command the reporting facility belongs.

**(4) Specify a Test Statistic**

$$\chi^2 = \sum_r \sum_c \frac{(O_{r,c} - E_{r,c})^2}{E_{r,c}}$$

where  $r$  and  $c$  run over the rows and columns of the observed and expected value matrices.

The P-value will be computed as

$$Pvalue = 1 - \text{chisq}(df, \chi^2)$$

with  $df$  defined as  $df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1)$

**(5) State the Distribution of the Test Statistic**

Under the null hypothesis, for a reasonable sample size, the statistic  $\chi^2$  computed above will approximately follow a  $\chi^2$  distribution having degrees of freedom

$$df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1)$$

**(6) Formulate a Decision Rule**

If  $Pvalue > \alpha$  then do not reject  $H_0$  in favor of  $H_a$  at significance level  $\alpha$

**(7) Obtain a Random Sample**

$$\text{Command} = \begin{pmatrix} 8 & 8 & 2 & 1 & 4 & 3 \\ 3 & 2 & 3 & 0 & 2 & 0 \end{pmatrix} \begin{matrix} \text{Reported} \\ \text{Did Not Report} \end{matrix}$$

The columns from left to right represent ACC, AFMC, AMC, PACAF, AETC and None.

**(8) Compute the Observed Value of the Test Statistic**

$$O = \text{Command}$$

$$r := 0 \dots \text{rows}(O) - 1 \quad c := 0 \dots \text{cols}(O) - 1$$

$$N := \sum_r \sum_c \text{Command}_{r,c} \quad N = 36$$

$$PC_c := \frac{\sum O^{<c>}}{N} \quad PC^T = (0.306 \quad 0.278 \quad 0.139 \quad 0.028 \quad 0.167 \quad 0.083)$$

$$\sum PC = 1$$

$$PR_r := \frac{\sum (O^T)^{<r>}}{N} \quad PR = \begin{pmatrix} 0.722 \\ 0.278 \end{pmatrix} \quad \sum PR = 1$$

Under the assumption of independence between column and row categories, the probability of being classified into the (r,c) cell will be

$$PR_r \cdot PC_c$$

the product of the corresponding row and column probabilities. A matrix of these probabilities can then be built...

$$P_{r,c} := PR_r \cdot PC_c \quad P = \begin{pmatrix} 0.221 & 0.201 & 0.1 & 0.02 & 0.12 & 0.06 \\ 0.085 & 0.077 & 0.039 & 0.008 & 0.046 & 0.023 \end{pmatrix} \sum_r \sum_c P_{r,c} = 1$$

The expected cell values are then given by  $E = N \cdot P$

$$E = \begin{pmatrix} 7.944 & 7.222 & 3.611 & 0.722 & 4.333 & 2.167 \\ 3.056 & 2.778 & 1.389 & 0.278 & 1.667 & 0.833 \end{pmatrix}$$

and the test statistic can then easily be computed

$$\chi^2 = \sum_r \sum_c \frac{(O_{r,c} - E_{r,c})^2}{E_{r,c}} \quad \chi^2 = 4.521$$

The degrees of freedom associated with this test statistic are:

$$df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1) \quad df = 5$$

The P-value is then calculated...

$$Pvalue := 1 - \text{chisq}(df, \chi^2) \quad Pvalue = 0.477 \quad \alpha := .1$$

**(9) Use the Decision Rule to state a Knowledge Claim**

$Pvalue > \alpha = 1$  indicates that  $Pvalue > \alpha$ , therefore do not reject the null hypothesis.

**(10) Consider the Practical Significance and State a Value Claim**

At a significance level of .1, the statistical analysis of the contingency table indicates that it is not possible to reject the claim that homogeneity exists across the data set.

**(11) Use the Knowledge Claim and the Value Claim to Answer the Focus Question**

The facility's status of reporting (Report/Did Not Report) does not appear to be related to which major command the reporting facility belongs.

**(1) State Focus Question to be Answered**

Is the facility's status of reporting (Report/Did Not Report) related to whether the major command the reporting facility belongs to is AFMC or some other command?

**(2) State Assumptions**

Assumptions:

1. Each observation is classified into exactly one cell.
2. The observations are observations on a random sample. Each observation has the same probability of being classified into cell (r,c) as any other observation.
3. The row totals are given, not random.

**(3) State Hypothesis**

$H_0$ : The populations of Report/Did Not Report are homogeneous with respect to whether the major command the reporting facility belongs to is AFMC or some other command?

$H_a$ : The populations of Report/Did Not Report are not homogeneous with respect to whether the major command the reporting facility belongs to is AFMC or some other command?

**(4) Specify a Test Statistic**

$$\chi^2 = \sum_r \sum_c \frac{(O_{r,c} - E_{r,c})^2}{E_{r,c}}$$

where  $r$  and  $c$  run over the rows and columns of the observed and expected value matrices.

The P-value will be computed as

$$Pvalue = 1 - \text{chisq}(df, \chi^2)$$

with df defined as  $df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1)$

**(5) State the Distribution of the Test Statistic**

Under the null hypothesis, for a reasonable sample size, the statistic  $\chi^2$  computed above will approximately follow a  $\chi^2$  distribution having degrees of freedom

$$df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1)$$

**(6) Formulate a Decision Rule**

If  $Pvalue > \alpha$  then do not reject  $H_0$  in favor of  $H_a$  at significance level  $\alpha$

**(7) Obtain a Random Sample**

$$\begin{array}{cc} \text{Yes} & \text{No} \\ \text{AFMC} = \begin{pmatrix} 8 & 18 \\ 2 & 8 \end{pmatrix} & \begin{array}{l} \text{Reported} \\ \text{Did Not Report} \end{array} \end{array}$$

**(8) Compute the Observed Value of the Test Statistic**

$$O = \text{AFMC}$$

$$r = 0 \dots \text{rows}(O) - 1 \quad c = 0 \dots \text{cols}(O) - 1$$

$$N = \sum_r \sum_c \text{AFMC}_{r,c} \quad N = 36$$

$$\text{PC}_c = \frac{\sum O^{<c>}}{N} \quad \text{PC}^T = (0.278 \quad 0.722) \quad \sum \text{PC} = 1$$

$$\text{PR}_r = \frac{\sum (O^T)^{<r>}}{N} \quad \text{PR} = \begin{pmatrix} 0.722 \\ 0.278 \end{pmatrix} \quad \sum \text{PR} = 1$$

Under the assumption of independence between column and row categories, the probability of being classified into the (r,c) cell will be

$$\text{PR}_r \cdot \text{PC}_c$$

the product of the corresponding row and column probabilities. A matrix of these probabilities can then be built...

$$P_{r,c} = \text{PR}_r \cdot \text{PC}_c \quad P = \begin{pmatrix} 0.201 & 0.522 \\ 0.077 & 0.201 \end{pmatrix} \quad \sum_r \sum_c P_{r,c} = 1$$



The expected cell values are then given by  $E = N \cdot P$

$$E = \begin{pmatrix} 7.222 & 18.778 \\ 2.778 & 7.222 \end{pmatrix}$$

and the test statistic can then easily be computed

$$\chi^2 = \sum_r \sum_c \frac{(O_{r,c} - E_{r,c})^2}{E_{r,c}} \quad \chi^2 = 0.418$$

The degrees of freedom associated with this test statistic are:

$$df = (\text{rows}(O) - 1) \cdot (\text{cols}(O) - 1) \quad df = 1$$

The P-value is then calculated...

$$Pvalue = 1 - \text{chisq}(df, \chi^2) \quad Pvalue = 0.518 \quad \alpha = .1$$

**(9) Use the Decision Rule to state a Knowledge Claim**

$Pvalue > \alpha = 1$  indicates that  $Pvalue > \alpha$ , therefore do not reject the null hypothesis.

**(10) Consider the Practical Significance and State a Value Claim**

At a significance level of .1, the statistical analysis of the contingency table indicates that it is not possible to reject the claim that homogeneity exists across the data set.

**(11) Use the Knowledge Claim and the Value Claim to Answer the Focus Question**

The facility's status of reporting (Report/Did Not Report) does not appear to be related to whether the major command the reporting facility belongs to is AFMC or some other command?

### Bibliography

"Air Force Material Command," Air Force Magazine, 78: 79 (May 1995).

Blackman, William C. Jr. Basic Hazardous Waste Management. Boca Raton, FL: Lewis Publishers, 1993.

Clinton, President William J. "Executive Order 12856: Federal Compliance with Right-to-Know Laws and Pollution Prevention Requirements". Federal Register, Vol. 58, No. 150. Washington: Government Printing Office, 6 August 1993.

Conover, W. J. Practical Nonparametric Statistics, Second Edition. New York: John Wiley & Sons, 1980.

Daniel, Wayne W. Biostatistics: A Foundation for Analysis in the Health Sciences, Fifth Ed. New York: John Wiley & Sons, 1991. p 548-549.

Department of Defense. Revised Implementing Guidance for Executive Order 12856: "Federal Compliance with Right-to-Know Laws and Pollution Prevention Requirements". Washington: DUSD(ES), 13 April 1995.

Devore, Jay L. Probability and Statistic for Engineering and the Sciences. Pacific Grove, CA: Brooks/Cole Publishing, 1991.

Dillman, Don A. Mail and Telephone Surveys: The Total Design Method. New York: John Wiley & Sons, 1978.

Environmental Protection Agency. Estimating Releases and Waste-Treatment Efficiencies for the Toxic Chemical Release Inventory Form. EPA 560/4-88-002. Washington: EPA Office of Pesticides and Toxic Substances, December 1987.

Environmental Protection Agency. Title III Section 313 Release Reporting Guidance: Estimating Chemical Releases From Semiconductor Manufacturing. EPA Document 560/4-88-004e. Washington: EPA Office of Pesticides and Toxic Substances, January 1988.

Environmental Protection Agency. The Emergency Planning and Community Right-to-Know Act: Section 313 Release Reporting Requirements. EPA 700/K-94-001. Washington: EPA Office of Pesticides and Toxics, April 1994a.

- Environmental Protection Agency. "Addition of Certain Chemicals; Toxic Chemical Release Reporting; Community Right-to-Know; Final Rule." Federal Register, Vol. 59, No. 229. Washington: Government Printing Office, 30 November 1994b.
- Environmental Protection Agency. 1993 Toxics Release Inventory: Public Data Release. EPA 745-R-95-010. Washington: Office of Pesticides and Toxics, March 1995a.
- Environmental Protection Agency. Guidance on Executive Order 12856: Federal Compliance with Right-to-Know Laws and Pollution Prevention Requirements. Washington: EPA, April 1995b.
- Fienberg, Stephen E. The Analysis of Cross-Classified Categorical Data. Cambridge, Massachusetts: MIT Press, 1978.
- GAIA Corporation. Emergency Planning and Community Right-to-Know Act Manual for U.S. Air Force Installations. Silver Springs, MD: GAIA Corporation, April 1994.
- GAIA Corporation. Emergency Planning and Community Right-to-Know Act Handbook for U.S. Air Force Installations. Silver Springs, MD: GAIA Corporation, February 1995.
- McPeak, General Merrill A., Air Force Chief of Staff. Environmental Leadership. Letter to All MAJCOMS, 17 April 1991.
- Nelson, Captain David K. The Pharmacy Concept for Hazardous Materials Management: Effect on Productivity of the Base Civil Engineer Operations Flight Work Force. MS thesis, AFIT/GEE/ENV/94S-16. School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, 1 August 1994a.
- Nelson, Captain David K. Environmental Program Manager, U.S. Air Force Headquarters, Washington D.C. Personal communication. 18 October 1994b.
- O'Brien, L.G. The Statistical Analysis of Contingency Table Designs. Norwich, England: Environmental Publications, University of East Anglia, 1989.
- Rice, Donald B. and General Merrill A. McPeak. Secretary of the Air Force and Air Force Chief of Staff. Air Force Pollution Prevention Program - Action Memorandum. Joint letter to All MAJCOMS, 7 January 1993.
- Snedecor, G.W. and W.G. Cochran. Statistical Methods, 7th ed. Ames, Iowa: Iowa State University Press, 1980.
- United States Air Force. Air Force Guidance for Toxic Release Inventory Reporting. Washington: USAF, 31 March 1995.

United States Air Force. DoD Implementing Guidance for Executive Order 12856.  
Washington: USAF, 28 February 1994.

### Vita

Captain Brian A. Pollock was born on 23 February 1964 in Springfield, Illinois. He graduated from Rochester Community High School in Rochester, Illinois in 1982 and attended Illinois State University for three semesters. He then enlisted in the United States Air Force, attended Basic Military Training at Lackland AFB, Texas, underwent technical training at Kessler AFB, Mississippi and was assigned to the 1500 Computer Services Squadron at Scott AFB, Illinois, as a computer operator in 1984. In 1985, he was selected for the Airman's Education and Commissioning Program and attended the University of Illinois, graduating with a Bachelor of Science degree in Industrial Engineering. Next, he attended Officer Training School and was commissioned a second lieutenant in the United States Air Force in May 1989. His first tour of duty was at Moody AFB, Georgia with the 347th Civil Engineering Squadron as Chief, Industrial Engineering. In 1991, he was reassigned to the 21st Civil Engineering Squadron, Elmendorf AFB, Alaska. His duties included Officer-In-Charge, Industrial Engineering, Squadron Reorganization Officer, and Deputy Chief, Maintenance Engineering. He was reassigned in May 1994 to the General Engineering and Environmental Management Master's Degree Program, School of Engineering, Air Force Institute of Technology. Captain Pollock's assignment following graduation will be with the 509th Civil Engineer Squadron at Whiteman AFB, Missouri.

Permanent Address: 1924 North 22nd Street  
Springfield, Illinois 62702

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.				
1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE December 1995		3. REPORT TYPE AND DATES COVERED Master's Thesis
4. TITLE AND SUBTITLE AN ANALYSIS OF 1994 AIR FORCE TOXIC RELEASE INVENTORY REPORTING				5. FUNDING NUMBERS
6. AUTHOR(S)  Brian A. Pollock, Capt, USAF				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)  Air Force Institute of Technology, WPAFB OH 45433-6583				8. PERFORMING ORGANIZATION REPORT NUMBER  AFIT/ENV/GEE/95D-19
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSORING / MONITORING AGENCY REPORT NUMBER
11. SUPPLEMENTARY NOTES				
12a. DISTRIBUTION / AVAILABILITY STATEMENT  Approved for public release; distribution unlimited				12b. DISTRIBUTION CODE
13. ABSTRACT (Maximum 200 words)  The purpose of this research is to analyze the Air Force's 1994 Toxic Release Inventory data. Statistical tests were used to meet the following research objectives: (1) review the CY 94 data to determine which chemicals were most often reported and which bases and major commands had the most releases in terms of the number of chemicals reported and the amount of chemicals reported and (2) investigate factors which could influence the reporting status of a facility. An analysis of the TRI data showed that dichloromethane was the chemical released in the largest amount while methyl ethyl ketone was the most often released chemical. The five Air Logistic Centers owned by Air Force Material Command, Tinker AFB, Kelly AFB, McClellan AFB, Hill AFB and Robins AFB, were responsible for the most TRI chemicals both in terms of numbers and pounds released. An analysis of a survey of Air Force facilities indicated that two factors had an influence on the reporting status of Air Force facilities: (1) whether the TRI point of contact had confidence in the completeness of the data used for threshold computations and (2) whether the primary database used for threshold computations was the Standard Base Supply System.				
14. SUBJECT TERMS  Toxic Release Inventory, Emissions, Toxic Chemicals, Environmental				15. NUMBER OF PAGES  124
				16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT  Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE  Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT  Unclassified	20. LIMITATION OF ABSTRACT  UL	